

Tree Physiology 38, 320–339
doi:10.1093/treephys/tpx054



Research paper

Genes and gene clusters related to genotype and drought-induced variation in saccharification potential, lignin content and wood anatomical traits in *Populus nigra*[†]

Henning Wildhagen^{1,10}, Shanty Paul¹, Mike Allwright², Hazel K. Smith², Marta Malinowska³, Sabine K. Schnabel⁴, M. João Paulo⁴, Federica Cattonaro⁵, Vera Vendramin⁵, Simone Scalabrin⁵, Dennis Janz¹, Cyril Douthe⁶, Oliver Brendel⁷, Cyril Buré⁷, David Cohen⁷, Irène Hummel⁷, Didier Le Thiec⁷, Fred van Eeuwijk⁴, Joost J.B. Keurentjes⁸, Jaume Flexas⁶, Michele Morgante⁹, Paul Robson³, Marie-Béatrice Bogeat-Triboulot⁷, Gail Taylor² and Andrea Polle^{1,11}

¹Forest Botany and Tree Physiology, Georg-August University of Göttingen, Büsingenweg 2, 37077 Göttingen, Germany; ²Center for Biological Sciences, University of Southampton, University Road, Southampton SO17 1BJ, UK; ³Institute of Biological, Environmental & Rural Sciences (IBERS), Aberystwyth University, Gogerddan, Aberystwyth, SY233EE, UK; ⁴Biométris, Wageningen University and Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands; ⁵IGA Technology Services, via Jacopo Linussio 51, 33100 Udine, Italy; ⁶Universidad de les Illes Balears, Carretera de Valldemossa Km 7.5, 07122 Palma de Mallorca, Illes Balears, Spain; ⁷EEF, INRA, Université de Lorraine, rue d'Amance, 54280 Champenoux, France; ⁸Laboratory of Genetics, Wageningen University and Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands; ⁹Università Di Udine, Istituto di Genomica Applicata, via Jacopo Linussio 51, 33100 Udine, Italy; ¹⁰Present address: HAWK University of Applied Sciences and Arts, Faculty of Resource Management, Büsingenweg 1a, 37077 Göttingen, Germany; ¹¹Corresponding author (apolle@gwdg.de)

Received January 17, 2017; accepted May 3, 2017; published online May 24, 2017; handling Editor Janice Cooke

Wood is a renewable resource that can be employed for the production of second generation biofuels by enzymatic saccharification and subsequent fermentation. Knowledge on how the saccharification potential is affected by genotype-related variation of wood traits and drought is scarce. Here, we used three *Populus nigra* L. genotypes from habitats differing in water availability to (i) investigate the relationships between wood anatomy, lignin content and saccharification and (ii) identify genes and co-expressed gene clusters related to genotype and drought-induced variation in wood traits and saccharification potential. The three poplar genotypes differed in wood anatomy, lignin content and saccharification potential. Drought resulted in reduced cambial activity, decreased vessel and fiber lumina, and increased the saccharification potential. The saccharification potential was unrelated to lignin content as well as to most wood anatomical traits. RNA sequencing of the developing xylem revealed that 1.5% of the analyzed genes were differentially expressed in response to drought, while 67% differed among the genotypes. Weighted gene correlation network analysis identified modules of co-expressed genes correlated with saccharification potential. These modules were enriched in gene ontology terms related to cell wall polysaccharide biosynthesis and modification and vesicle transport, but not to lignin biosynthesis. Among the most strongly saccharification-correlated genes, those with regulatory functions, especially kinases, were prominent. We further identified transcription factors whose transcript abundances differed among genotypes, and which were co-regulated with genes for biosynthesis and modifications of hemicelluloses and pectin. Overall, our study suggests that the regulation of pectin and hemicellulose metabolism is a promising target for improving wood quality of second generation bioenergy crops. The causal relationship of the identified genes and pathways with saccharification potential needs to be validated in further experiments.

Keywords: biofuels, cell wall, drought, genotypic variation, glucose release, lignin, *Populus*, saccharification, wood traits.

[†]Dedicated to the memory of Carl Douglas.

© The Author 2017. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution and reproduction in any medium, provided the original work is properly cited. For commercial

Introduction

Wood is an attractive feedstock for biofuels because it is a renewable resource that can be produced in a sustainable manner (Polle et al. 2013, Weih and Polle 2016). Wood is composed mainly of cellulose (35–50%), hemicelluloses (15–35%), lignin (15–35%) and pectins (<10%) (Plomion et al. 2001). Glucose can be released from cellulosic compounds in the cell wall by saccharification and subsequently used for conversion into ethanol (Galbe and Zacchi 2002). Lignin and hemicelluloses and their interactions with pectin or cellulose can negatively impact the efficiency of biofuel production (Mansfield et al. 1999, Nookaraju et al. 2013, Xiao and Anderson 2013). For example, transgenic lines of *Populus trichocarpa* with suppressed expression of 4-coumarate:coenzyme A ligase contained lower lignin contents and exhibited greater saccharification potential than the wild type (Min et al. 2012). In *Populus × canescens*, in which cinnamoyl-CoA reductase, a key enzyme for lignin biosynthesis, was down-regulated, the cell walls were more easily digestible, which resulted in higher saccharification and ethanol yield (Acker et al. 2014).

The saccharification potential of wood is also affected by environmental conditions such as drought (Fiasconaro et al. 2012, van der Weijde et al. 2016). Drought brings about changes in cell wall composition including an increased abundance of hemicellulose, pectins and lignin, which results in strengthening of the cell wall (Moore et al. 2008, Le Gall et al. 2015). Fiasconaro et al. (2012) reported that the potential for biofuel conversion in drought-treated alfalfa plants decreased due to a higher lignin content compared with the well-watered plants. On the other hand, in a study comprising 50 different *Miscanthus* genotypes, drought treatment caused a reduction of cell wall cellulose content and an increase in cell wall hemicellulose content, but nevertheless resulted in a higher efficiency of the conversion of cellulose to glucose during saccharification (van der Weijde et al. 2016). These contrasting findings suggest that the effect of drought on saccharification potential depends largely on the particular species studied, and on the relative changes of cell wall components under drought. In poplar, drought stress has massive consequences for the wood anatomy and cell wall metabolism (Harvey and Van Den Driessche 1997, Arend and Fromm 2007, Beniwal et al. 2010, Fichot et al. 2009, 2010, Schreiber et al. 2011, Cao et al. 2014, Guet et al. 2015, Le Gall et al. 2015), but studies on the effects of drought on the saccharification potential of *Populus* and the underlying anatomical and molecular responses have not yet been reported, but are greatly needed (Studer et al. 2011).

The saccharification potential is also subject to genetic variation, as shown for *Miscanthus* (Souza et al. 2015) and poplar (Studer et al. 2011). Similarly, natural genetic variation in lignin content was reported for many taxa, including *Arabidopsis thaliana* (Capron et al. 2013), eucalyptus (Klash et al. 2010, Elissetche et al. 2011), conifers (Gonzalez-Martinez et al. 2007,

Gaspar et al. 2011) and poplars (Wegrzyn et al. 2010, Zhou et al. 2011, Guerra et al. 2013, Porth et al. 2013, Muchero et al. 2015). Given the importance of the genus *Populus* as a second generation bioenergy crop (Allwright and Taylor 2016), it is of great interest to study whether these variations translate into variation in saccharification potential.

Attempts to improve the saccharification potential of bioenergy crops require insights into the molecular control of wood properties. Several studies using quantitative trait locus (QTL) mapping or association mapping approaches identified candidate genes related to cell wall composition, as well as lignin content and composition (Ranjan et al. 2009, Guerra et al. 2013, Porth et al. 2013, Fahrenkrog et al. 2016). However, such approaches have only rarely been applied to uncover candidate genes or QTLs related to sugar release (Brereton et al. 2010, Muchero et al. 2015). Knowledge on transcriptional networks controlling diverse aspects of wood formation including secondary cell wall formation and biosynthesis of its components has accumulated during recent years (for recent reviews: Zhong et al., 2010, Hussey et al., 2013, Nakano et al., 2015, Ye and Zhong 2015). However, a systematic transcriptome-wide investigation of genes and gene clusters underlying genetic variation in saccharification potential in economically important bioenergy crops such as *Populus* is missing.

The aim of this study was to identify clusters of co-expressed genes, as well as candidate genes and their putative transcriptional regulators, related to genotype- and drought-induced variation in wood anatomy, lignin content and saccharification potential. We hypothesized (i) that drought results in increased lignification and decreased saccharification, (ii) that genotypes originating from different environments differ in lignin content and saccharification potential and (iii) that these drought and genotype-effects on saccharification potential are underpinned by distinct differences in wood anatomical traits and transcript abundances of genes involved in wood formation and cell wall metabolism, especially lignin biosynthesis.

To this end, we used three genotypes that were selected from large-scale common garden experiments with up to 13 different *Populus nigra* L. populations (DeWoody et al. 2015, Viger et al. 2016). The *P. nigra* populations originated from habitats with different climatic conditions across Europe and showed significant genetic differentiation as well as phenotypic variation in growth rates, plant architecture and leaf size (DeWoody et al. 2015, Viger et al. 2016). The most pronounced differences among the phenotypic traits were observed between *P. nigra* genotypes originating from a dry habitat in Spain and those from a wet habitat in Italy, while a French genotype exhibited intermediate properties (Viger et al. 2016). Here, we exposed these contrasting genotypes to a highly controlled drought treatment, applied RNA sequencing and weighted gene correlation network analysis (WGCNA) to identify novel candidate genes related to wood properties and saccharification potential.

Materials and methods

Plant material

Three genotypes of *P. nigra* L., originating from individual trees of natural populations in France (Drôme 6; FR-6), Italy (La Zelata; IT1) and Spain (Ebro 2; SP-2; see DeWoody et al. 2015) were studied. Cuttings from these genotypes were purchased from INRA Orleans (Orléans, France), where the materials are available for research purposes under the accession numbers N-38 ('Italy'), RIN2-new ('Spain') and 6-J29 ('France'). The cuttings were planted in 10 l plastic pots filled with a 1:1 (v/v) mixture of peat and sand, amended with a slow-release fertilizer (4 g l⁻¹ of Nutricote T100, 13:13:13 NPK and micronutrients; FERTIL S.A.S., Boulogne Billancourt, France) and 1 g l⁻¹ CaMg(CO₃)₂. Plants were grown in two chambers of a fully automated glasshouse for phenotyping located at Champenoux, France (48°45'09.3"N, 6°20'27.6"E), under natural light conditions with daily maxima of irradiance ranging from 154 to 1011 μmol m⁻² s⁻¹ photosynthetically active radiation. Environmental conditions in the greenhouse were affected by weather conditions, but the temperature was adjusted to not exceed 28 °C. Plants were watered three times a day to 85% of field capacity with an automated watering system.

Drought experiment

After 6 weeks of growth, 32 plants of each genotype were randomly assigned to either a control or a drought treatment. The plants were randomized across the two greenhouse chambers, so that each chamber contained balanced proportions of each genotype and treatment. The position of plants was changed automatically.

Plants were exposed to drought by gradually decreasing the available soil water content (SWC). The control of SWC was based on pot weight and a calibration between volumetric SWC measured by Time Domain Reflectometry and pot weight. Available water was then expressed as relative extractable water content (REW_{soil}), which is defined as:

$$REW_{soil} = \left(\frac{SWC - SWC_{wilting\ point}}{SWC_{field\ capacity} - SWC_{wilting\ point}} \right) \times 100\%,$$

with SWC at wilting point = 3%; SWC at field capacity = 32%. Sixteen control plants per genotype were watered to 85% REW_{soil} three times a day by the automated system for the whole 5-week period of the experiment. For drought-treated plants, REW_{soil} was progressively decreased to reach a target level of 20% REW_{soil} in 2 weeks. The decrease in REW_{soil} was controlled and adjusted in each individual pot because bigger plants used more water than smaller plants. Thereby, all sixteen drought-treated plants per genotype experienced the same decline in REW_{soil}. Plants were then watered by the automated system to keep REW_{soil} at this target level for the following 3 weeks.

After 5 weeks of control or drought treatment, all plants were harvested. The developing secondary xylem was collected as described in Teichmann et al. (2008) from 10 plants per genotype and treatment, frozen immediately in liquid nitrogen and stored at -80 °C until further analysis. For wood anatomical analyses, the bottommost straight upright segment of 2–3 cm length of the stem was collected and immediately fixed in FAE (2% formaldehyde, 5% acetic acid and 63% ethanol). For the analysis of saccharification potential and lignin content, a stem segment of 3–4 cm length above the piece of wood collected for anatomical analysis was debarked and dried to constant weight at 50 °C.

Plant radial growth

The diameter of the stem was measured twice a week on six biological replicates by taking photographs of the stem base with a scale bar attached to a fixed position. Stem diameter was then determined by image analysis using the software ImageJ (Schneider et al. 2012). Based on these measurements, the daily radial area increment A_{inc} for the period when target level of drought was reached (Day 13) until final harvest (Day 36) of the experiment was calculated as follows:

$$A_{inc} = \frac{\pi r_{Day\ 36}^2 - \pi r_{Day\ 13}^2}{Day\ 36 - Day\ 13}.$$

Wood anatomical analyses

FAE-fixed stem pieces were transferred into water to remove FAE. Stem cross sections of 10 μm thickness were cut with a freezing microtome (Jung AG, Heidelberg, Germany). The cross-sections were transferred into water and stained in 0.05% (w/v, pH = 7.0) toluidine blue O solution (O'Brien et al. 1964) for 3 min. The sections were then washed and mounted on glass slides using glycerol (28% v/v) as mounting medium. The prepared sections were viewed under a microscope (Axioplan Observer.Z1, Carl Zeiss GmbH, Oberkochen, Germany). Images were taken with a digital camera (Axio Cam MRC, Carl Zeiss Microimaging GmbH, Göttingen, Germany) attached to the microscope.

Images were analyzed with ImageJ (Schneider et al. 2012). Areas of 300 μm × 300 μm or of 100 μm × 200 μm in the mature secondary xylem adjacent to the developing xylem region were used for trait measurements at 100× (vessel frequency, vessel lumen area) or 400× (vessel cell wall thickness, fiber double wall thickness, fiber frequency, fiber lumen area, percentage of cell wall area of vessels and fibers, percentage of ray area, number of cambial cell layers) magnifications, respectively. These areas were selected because all cells here were newly formed during the treatment period. Five biological replicates per genotype and treatment were analyzed.

Five measurements of vessel wall thickness were taken for each sample. To determine fiber double wall thickness, 14 measurements were made for each sample. Vessel or fiber frequency was defined as number of vessels or fibers per unit area. The number

of vessels or fibers was calculated per mm^{-2} . The lumen area of each vessel in the analyzed xylem area ($300 \mu\text{m} \times 300 \mu\text{m}$) of a sample was measured using the 'Analyze particles' option of ImageJ. Fiber lumen area was measured manually with the help of 'freehand selection tool' in ImageJ. To obtain total fiber lumen area of a specified area, the average fiber lumen area obtained for a subsample of 14 fibers was multiplied with the total fiber number. The fraction of cell wall area (CWA) per total area (%) of analyzed xylem (TAX) of each sample was calculated as follows:

$$\text{CWA}(\%) = \frac{[\text{TAX} - (\text{vessel lumen area} + \text{fibre lumen area} + \text{ray area})] \times 100}{\text{TAX}}$$

For calculating relative width of developing xylem, first the radius of the developing xylem ring was measured using images taken at 100 \times magnification while the radius of the whole xylem (developing and mature secondary xylem) was calculated from images taken at 25 \times magnification as the distance from the cambium to the pith. The relative width of the developing xylem was calculated as:

Relative width of developing xylem (%) = (radial width of developing xylem / radial width of whole xylem) \times 100.

Analysis of lignin content

Cell walls were isolated adapting the procedure by Foster et al. (2010) and Da Costa et al. (2014). For each 70 mg of plant material, cell walls were extracted following this sequence of steps: the material was incubated twice with 1.5 ml of aqueous ethanol for 20 min at 40 °C; once with 1.5 ml chloroform/methanol (1:1 v/v) for 10 min; and three times for 10 min with 1.5 ml acetone. During each incubation, samples were vortexed frequently. Between each step, the supernatant was collected by centrifugation at 15,000 r.p.m., and discarded. Following the last acetone wash, samples were left to dry at 35 °C overnight in a fume hood. Dried biomass was then re-suspended in 1.5 ml of 0.1 M sodium acetate buffer (pH 5.0) and heated to 80 °C for 20 min to induce starch gelatinization. Once cooled, 35 μl of 0.01% sodium azide was added to inhibit microbial growth, and starch was removed by incubation with 35 μl amylase (50 $\mu\text{g}/\text{ml}$ in H_2O ; from *Bacillus* sp., (Sigma, St Louis, MO, USA)) and 17 μl Pullulanase (18.7 units from *Bacillus acidopullulyticus*; Sigma). Tubes were incubated overnight at 37 °C and shaking at 400 r.p.m. The next day the digestion was terminated by heating the samples at 98 °C for 10 min. The cell wall material was collected by centrifugation. Purified cell walls were then washed three times with 1.5 ml deionized water and twice with 1.5 ml of acetone followed by drying in the fume hood.

For each sample, acetyl bromide soluble lignin was determined in triplicate following the procedures described in Da Costa et al. (2014) and Foster et al. (2010). Aliquots of 7.0 mg (± 0.3) of the cell-wall samples were weighed into 10 ml Pyrex glass tubes. To solubilize lignin, 500 μl of the freshly prepared acetyl bromide solution (25% v/v acetyl bromide in glacial acetic acid) was added.

Capped tubes were incubated for 3 h at 50 °C; during the last 30 min of incubation, samples were mixed three times using a vortex mixer. After the digestion tubes were cooled on ice and the samples were diluted by the addition of 2000 μl of 2 M NaOH and 350 μl 0.5 M freshly prepared hydroxylamine hydrochloride and gently mixed. The solution was then neutralized with 7150 μl glacial acetic acid. The tubes were recapped, and mixed several times by inversion and then centrifuged at 1000 r.p.m. for 5 min to pellet undissolved cell wall material. An aliquot of 200 μl of each sample in triplicate was transferred to UV-transparent 96-well plates (UV-Star; Greiner Bio-One, Gloucestershire, UK). The absorbance of each assay mixture was determined two to three times at 280 nm using a plate reader (mQuant; Bio-Tek Instruments, Winooski, VT, USA) and KC4 software (v. 3.3; Bio-Tek). Included in each plate was a standard cell wall preparation of poplar, which was a large mixed sample of cell wall extract that had been assayed 20 times to provide an accurate assay value to test for consistency in subsequent assays; also included was one negative control as the absorbance baseline, which contained all reactants but lacked cell wall material. Acetyl bromide-soluble lignin percentage content (ABSL%) was calculated using absorption reading at 280 nm (A_{280}), path length determined for the 96-well microplates with a volume of 200 μl per well used during the analysis (0.556 cm) (PL), reaction volume in liters (VR) the sample weight in grams (WS), and the specific absorption coefficient (SAC, $17.78 \text{ g}^{-1} \text{ l}^{-1} \text{ cm}^{-1}$) in the equation described by Da Costa et al. (2014):

$$\text{ABSL} = \left(\frac{A_{280}}{\text{SAC} \times \text{PL}} \right) \times \frac{\text{VR}}{\text{WS}} \times 100\%.$$

Analysis of saccharification potential

Samples were assayed for saccharification potential according to the methodology described by Van Acker et al. (2013). Samples were dried overnight at 50 °C and milled for 1 min in a Retsch 300MM Mixer Miller (Retsch GmbH, Haan, Germany) at a frequency of 25 s^{-1} . Ground material was sieved and the fraction falling between 150 and 850 μm retained. An aliquot of 10–11 mg of ground material was used for each assay. Samples received acid pretreatment for 2 h at 80 °C in 1 M HCl and washing in 70% ethanol for 20 h at 55 °C. Samples were rinsed with water and acetone prior to re-drying overnight at 50 °C. Samples were reweighed before undergoing 48 h saccharification at 55 °C in a rotating thermomixer in acetic acid buffer with fungal cellulase (*Trichoderma reesei*) and cellobiase (*Aspergillus niger*) enzymes (Sigma) with 0.06 filter paper units activity (Xiao et al. 2004). Supernatant post-saccharification was assayed with GOD-POD (glucose oxidase, horseradish peroxidase and ABTS dye) solution permitting spectrophotometric (ELx800 Absorbance Reader, BioTek) glucose quantification from sample absorbance at 405 nm. Saccharification potential was calculated as sample glucose yield as a percentage of post-pretreatment, dry cell wall residue (PPT CWR).

Statistical analysis of growth and wood anatomical data

Statistical analyses were performed using R v3.1.1 (R Development Core Team 2015). Two-factorial mixed linear models with 'genotype' and 'treatment' included as main and interaction effect as well as a random effect for greenhouse chamber were fitted to the data using the function 'lme', package 'nlme' (Pinheiro et al. 2015). Normality and homogeneity of variance were tested visually by plotting the residuals. The data were transformed logarithmically (log₂) if needed to achieve normality and homogeneity of variance of the residuals. In case of a significant ($P < 0.05$) genotype main effect, post-hoc tests were computed using the function 'lsm' (package 'lsmeans'; Lenth 2015) called within the function 'ghlt' (package 'multcomp'; Hothorn et al. 2008).

RNA extraction, library preparation and sequencing

Total RNA was extracted from homogenized samples of developing xylem of four biological replicates per genotype and treatment using the CTAB protocol (Chang et al. 1993) with minor modifications described in Janz et al. (2012). Isolated RNA was quality checked using Agilent 2100 Bioanalyzer RNA Nano assay (Agilent Technologies, Santa Clara, CA, USA). Two micrograms of total RNA (RNA Integrity Number RIN >7.0) was used for library preparation using the 'TruSeq mRNA Sample Prep kit v2' (Illumina, San Diego, CA, USA), following the manufacturers' instructions. Final libraries were quantified using the Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) and quality tested by Agilent 2100 Bioanalyzer High Sensitivity or DNA 1000 assay (Agilent Technologies). Successively, libraries were loaded on Illumina cBot for cluster generation on the flow cell, following the manufacturer's instructions and sequenced in 50 bp single-end mode at sixfold multiplex on the Illumina HiSeq2000 (Illumina). The CASAVA 1.8.2 version of the Illumina pipeline was used to process raw data for both format conversion and de-multiplexing. In average, 26.91 million reads were produced per sample (min 21.12). Short reads have been deposited into NCBI Short Read Archive under BioProject accession PRJNA359401.

Bioinformatic analyses

Raw sequence reads were processed with the Python package Cutadapt v1.4.2 (Martin 2011) to remove any residual adapter contamination. Reads were subsequently trimmed to remove low-quality reads (option -trim_qual_left/right = 25) and reads shorter than 40 nucleotides, using the PRINSEQ software v.lite-0.20.4 (Schmieder and Edwards 2011). Trimmed reads were aligned to the *P. trichocarpa* v3.0 transcriptome database (Tuskan et al. 2006) using TopHat2 v2.0.10 (Kim et al. 2013). A count table was generated using the Python package HTSeq v0.6.1 (Anders et al. 2015). Further inferential analyses of the count data were done with the R package DESeq2, version 1.6.3 (Love et al. 2014) unless otherwise stated. Normalization of count tables was done based on the 'median ratio method' (Anders and Huber 2010) implemented in

the function 'estimateSizeFactors'. The analyzed libraries were part of a larger set of libraries generated from various plant tissues. Prior to any down-stream analyses, we identified genes significantly affected by a greenhouse effect across all tissues. By this procedure, three genes were identified and removed from the initial count table.

Prior to the analysis of differential expression, we applied an unspecific filtering (Bourgon et al. 2010) to keep only those genes to which at least one read per million reads of library size aligned in at least four samples. The analysis of differential expression was carried out by fitting two-factorial generalized linear models of the negative binomial family (NB) (function 'DESeq') to the read counts K_{ij} for gene i in sample j with a logarithmic link (Love et al. 2014):

$$K_{ij} \sim \text{NB}(\text{mean} = \mu_{ij}, \text{dispersion} = \alpha_i)$$

$$\mu_{ij} = s_j q_{ij}$$

$$\log q_{ij} = \sum_r x_{jr} \beta_{ir}$$

The computations of the normalization constant s_j and the dispersion parameter α_i are detailed in Love et al. (2014); q_{ij} is proportional to the expectation value of the true concentration of fragments from gene i in sample j , x_{jr} denotes the elements of the design matrix X , and β_{ir} denote the coefficients for gene i and parameters (corresponding to columns of the design matrix) r .

To test for the significance of the 'genotype-by-treatment' interaction effect, the full model including 'genotype' and 'treatment' main and interaction effects (design matrix X_1 in Table S1 available as Supplementary Data at *Tree Physiology* Online) was compared against a reduced model without interaction effect (design matrix X_2 in Table S1 available as Supplementary Data at *Tree Physiology* Online) by applying a likelihood ratio test implemented in the function 'nbinomLRT'. To assess the significance of the 'treatment' and 'genotype' main effect, a full model without interaction effect was set up (design matrix X_3 in Table S1 available as Supplementary Data at *Tree Physiology* Online), against which the respective reduced models without 'treatment' (design matrix X_4 in Table S1 available as Supplementary Data at *Tree Physiology* Online) or 'genotype' main effect (design matrix X_5 in Table S1 available as Supplementary Data at *Tree Physiology* Online) were tested with the function 'nbinomLRT'. 'Treatment' referred to drought treatment as main effect; genes with significant differences (false discovery rate [FDR] = 0.05) in transcript abundances between drought-treated and control tissue were denominated as drought-related differentially expressed genes (DDEGs). Genes with a significant genotype main effect (FDR = 0.05) in transcript abundance were denominated as genotype-related differentially expressed genes (GDEGs). Pairwise genotype contrasts were computed for the set of GDEGs by specifying the contrasts in the function 'results'.

Gene co-expression analysis and correlation to wood traits

The GDEGs and DDEGs, which had been obtained by the bioinformatics analyses described above, were used to re-construct transcriptional networks using the R package WGCNA version 1.47 (Langfelder and Horvath 2008, 2012). Gene co-expressions were summarized as adjacency matrices using soft-thresholding powers of 20. The soft-thresholding powers were selected based on low mean connectivity and a >90% model fit to scale-free topology. Adjacency matrices were transformed into topological overlap matrices, which were used to determine modules of co-expressed genes, with a cut height of 0.99 and the minimal module size set to 20 and 30 for the network of DDEGs and GDEGs, respectively. Modules whose correlation of eigengene values was >0.8 were merged. Eigengene values (corresponding to the first principal component of the modules' gene expression data) of modules of DDEGs or GDEGs were then correlated with wood traits.

Enrichment of gene ontology (GO) categories (Ashburner et al. 2000) within modules was tested using 'The Ontologizer' (Bauer et al. 2008). Genes assigned to individual modules were used as study sets, and the sets of all genes showing a drought- or genotype-main effect, respectively, as the reference sets ('gene population').

The transcript abundances of the individual genes in modules related to the saccharification potential were tested for significant correlation with the saccharification potential. The resulting set was mapped onto MapMan categories (MapMan v3.5.1R2; Thimm et al. 2004) based on their *A. thaliana* best hits. MapMan's built-in Wilcoxon rank sum test with Benjamini-Yekutieli correction was used to identify bins enriched in genes with high correlations to saccharification potential. *Arabidopsis* best hits assigned to the MapMan bin 'cell wall' (bin 10) were used as baits in the co-expression database CressExpress (Srinivasan et al. 2008; <http://cressexpress.org/>) to identify genes commonly co-regulated with these bait genes across multiple experiments. CressExpress v3.0 was used with default quality-control statistic D to search for co-regulated genes in the microarray data from experiments NASC_137 ('AtGenExpress Stress Treatments, (control plants); GSE5620; Kilian et al. 2007), NASC_141 ('AtGenExpress Stress Treatments' (drought stress); GSE5624; Kilian et al. 2007), NASC_153 ('AtGenExpress Developmental Series shoots and stems'; GSE5633), NASC_14 ('control of lignification'; Rogers et al. 2005) and NASC_27 ('Assembly of the cell wall pectic matrix', GSE6181). Genes were considered co-expressed with the saccharification-related bait genes when the gene-expression data across the selected microarray experiments were correlated with $r^2 > 0.8$ and $P < 0.05$. Co-expressed genes annotated as transcription factors were identified according to the list of *A. thaliana* transcription factors available at the Plant Transcription Factor Database v4.0 (Jin et al. 2016; <http://plantfdb.cbi.pku.edu.cn/>). The identified transcription factors were

compared with the transcription factors in the list of DEGs and genes present in both lists were extracted as candidate genes.

Quantitative real time polymerase chain reaction (qRT-PCR) of selected genes

RNAseq data were validated by quantitative real time polymerase chain reaction (qRT-PCR) on a Light Cycler 480 system (Roche Diagnostics, Mannheim, Germany). Primers were designed with Oligo Explorer (Gene Link, Hawthorne, NY, USA, <http://www.genelink.com/>). We used three genes without significant changes across our conditions as the references (Potri.012G141400, Potri.001G006400, Potri.005G074900; primer sequences (5'–3', forward and reverse): AAATTGAGGTTGGGAAGGC, ACAACATC TGGCATGCATCC for Potri.012G141400; GATCCTCATGATGCTG TGA, GCAACTGTACGCTCCCTTG for Potri.001G006400; ACCAA CGAGACAAGGTGCTT, CTTTTGGGCTTCTTGCAAAC for Potri.005G074900), and four genes as targets (Potri.014G035500, Potri.012G097900, Potri.010G083600, Potri.004G038700; primer sequences (5'–3', forward and reverse): TTCGAGCCC AACATTACAAAG, GATAGTAGGTGGGGATGGCATG for Potri.014G035500; AATCTCCCTCATTGCCATCAC, ATGATGATCAGAC GCCGCTGG for Potri.012G097900; GGTTGCAGGAATTAGGAAC GC, ACGTTTCCCAGTTCCATGTTG for Potri.010G083600; CGGTTC ACTCCTCTTCTTATG, TGGCAAGAGTGAGAAGGATCTG for Potri.004G038700). Primer efficiencies and C_q-values were obtained using LinRegPCR v2016.2.0.0 (Ruijter et al. 2009). The expression of target genes was related to the expression of the reference genes as the inverse of the normalized relative quantity defined in Hellemans et al. (2007). The correlation of the relative expression of target genes determined by qRT-PCR with normalized transcript abundance obtained by RNAseq was analyzed. The following correlations were obtained: Potri.014G035500: $r = 0.63$, $P = 0.001$, Potri.012G097900: $r = 0.73$, $P < 0.001$, Potri.010G083600: $r = 0.79$, $P < 0.001$, Potri.004G038700: $r = 0.38$, $P = 0.071$ (Figure S1 available as Supplementary Data at *Tree Physiology* Online).

Results

Genotypic and drought-induced variation in wood anatomical traits

The three *P. nigra* genotypes originating from habitats with increasing precipitation (Spain < France < Italy; DeWoody et al. 2015), differed significantly in radial area increment ($P < 0.0001$). Genotype France showed a significantly higher radial area increment as compared with the genotypes Italy and Spain (Table 1). Drought resulted in a significant decline in radial growth ($P < 0.0001$), while the genotype–drought interaction effect was not significant ($P = 0.662$).

In order to elucidate whether the variation in radial area production was related to changes in cambial activity or the

differentiation of newly formed xylem cells, we counted the cambial cell layers and determined the radial width of the developing xylem. The number of cambial cell layers did not vary among genotypes (Table 1). Significant genotypic variation was observed in the relative radial width of the developing xylem ($P = 0.014$), with genotype Italy having a smaller relative width of the developing xylem compared with genotype France (Table 1). The genotypes differed significantly in vessel ($P = 0.007$) and fiber frequencies ($P = 0.02$) as well as in the vessel ($P = 0.007$) and fiber lumen areas ($P < 0.0001$, Table 1). The genotype France showed lower vessel and fiber frequencies and significantly higher vessel and fiber lumen areas compared with the other two genotypes (Table 1). Neither vessel nor fiber cell

wall thickness or cell wall area of vessels and fibers differed among the *P. nigra* genotypes studied here (Table 1). Genotype Spain showed a significantly higher percentage of ray area as compared with genotype France ($P = 0.034$). Under drought, both the number of cambial cell layers and the relative radial width of the developing xylem decreased (Table 1), reflecting the significant reduction in radial area growth under drought ($P < 0.0001$). The vessel frequency increased, while the decrease in vessel lumen area under drought was marginally significant ($P = 0.062$, Table 1). Among the fiber traits, only the fiber lumen area was reduced under drought ($P = 0.023$), whereas the fiber frequency was unaffected in the three genotypes (Table 1). The fraction of cell wall area of vessels, fibers and rays

Table 1. Wood anatomical traits and lignin of three genotypes of *Populus nigra* exposed to a control or drought treatment for 5 weeks.

Anatomical trait	Genotype	Mean \pm SE		P-value			Post-hoc
		Control	Drought	Drought	Genotype	G \times D	
Radial area growth ($\text{mm}^2 \text{day}^{-1}$)	Spain	1.9 (± 0.12)	0.9 (± 0.12)	<0.0001	<0.0001	0.663	a
	France	2.9 (± 0.04)	1.9 (± 0.13)				b
	Italy	2.1 (± 0.12)	1.3 (± 0.09)				a
Relative radial width of developing xylem (%)	Spain	5.3 (± 0.63)	4.0 (± 0.40)	0.005	0.014	0.294	ab
	France	5.3 (± 0.54)	4.8 (± 0.57)				b
	Italy	4.5 (± 0.27)	2.9 (± 0.24)				a
Number of cambial cell layers	Spain	6.4 (± 0.2)	4.8 (± 0.4)	<0.0001	0.659	0.241	a
	France	7.0 (± 0.3)	4.4 (± 0.4)				a
	Italy	7.2 (± 0.4)	4.6 (± 0.25)				a
Vessel frequency (vessel number mm^{-2})	Spain	170 (± 12)	199 (± 13)	0.007	0.007	0.704	ab
	France	124 (± 12)	163 (± 5)				a
	Italy	176 (± 16)	264 (± 59)				b
Vessel lumen area per vessel (μm^2)	Spain	1285 (± 61)	1075 (± 68)	0.062	0.007	0.720	a
	France	1643 (± 137)	1371 (± 125)				b
	Italy	1173 (± 42)	1048 (± 213)				a
Vessel wall thickness (μm)	Spain	1.10 (± 0.06)	1.15 (± 0.08)	0.222	0.139	0.841	a
	France	1.26 (± 0.09)	1.30 (± 0.1)				a
	Italy	1.16 (± 0.07)	1.28 (± 0.05)				a
Fiber frequency (fiber number mm^{-2})	Spain	3894 (± 210)	4383 (± 290)	0.595	0.02	0.230	ab
	France	3925 (± 248)	3607 (± 314)				a
	Italy	5003 (± 431)	4411 (± 293)				b
Fiber lumen area per fiber (μm^2)	Spain	98 (± 11)	82 (± 3.7)	0.023	<0.0001	0.235	a
	France	125 (± 9)	124 (± 5.7)				b
	Italy	97 (± 7)	71 (± 2.5)				a
Fiber double wall thickness (μm)	Spain	3.8 (± 0.16)	3.9 (± 0.19)	0.495	0.464	0.998	a
	France	3.6 (± 0.23)	3.7 (± 0.27)				a
	Italy	3.5 (± 0.19)	3.7 (± 0.24)				a
Cell wall area of vessels and fibers (%)	Spain	33.8 (± 2.8)	36.8 (± 2.2)	0.023	0.13	0.828	a
	France	27.8 (± 0.1)	33.1 (± 3.0)				a
	Italy	31.0 (± 2.6)	36.7 (± 1.5)				a
Percentage of ray area (%)	Spain	9.9 (± 1.7)	9.5 (± 1.0)	0.803	0.025	0.988	b
	France	6.8 (± 1.1)	6.5 (± 1.1)				a
	Italy	7.0 (± 1.2)	6.9 (± 0.5)				ab
Lignin content per dry well wall (%)	Spain	17.2 (± 0.56)	16.9 (± 0.40)	0.269	0.003	0.348	ab
	France	15.6 (± 0.49)	16.1 (± 0.51)				a
	Italy	17.1 (± 0.57)	18.2 (± 0.40)				b

Mean \pm SE are given for $n = 5$ – 6 biological replicates per genotype and treatment. P -values as computed for 2-factorial linear models including genotype and drought main effects and the genotype–drought interaction effect (G \times D). Significant effects with $P < 0.05$ are indicated in bold letters. Post hoc: letters indicate homogenous subgroups of genotypes identified by post-hoc tests for traits showing a significant genotype effect.

was significantly larger in drought-treated than in control trees ($P = 0.023$). There was no significant genotype–drought interaction effect in any of the analyzed wood anatomical traits.

Saccharification potential and lignin in *P. nigra* genotypes in relation to wood anatomical traits

The lignin content (per cell wall mass) showed significant genotypic variation, with lower values in genotype France than in genotype Italy ($P = 0.003$), but was unaffected by drought (Table 1). Saccharification potential was significantly higher in genotype Spain compared with genotypes France and Italy (Figure 1), and increased in response to drought ($P = 0.021$, Figure 1).

We tested whether relationships existed among the chemical and wood anatomical traits (Figure 2). The anticipated negative relationship between lignin and saccharification potential was not observed. The saccharification potential showed only a weak relationship with fiber lumen ($P = 0.047$, Figure 2). In contrast to the saccharification potential, wood traits and lignin were connected by various correlations. For example, the lignin content was negatively correlated with vessel ($P = 0.026$) and fiber lumina ($P = 0.005$), suggesting links between cell expansion and lignification. The strongest positive correlations were found between double wall thickness of fibers and the fraction of cell wall area ($P < 0.001$) and between fiber lumen area and relative width of developing xylem ($P < 0.001$, Figure 2) indicating that fiber expansion may be a stronger driver of growth than vessel expansion ($r = 0.63$ for fiber lumen area compared with $r = 0.49$ for vessel lumen area) and that thick fiber cell walls lead to a high fraction of cell wall area, i.e., dense wood (Figure 2). The strongest negative relationships were found between vessel frequency and vessel lumen ($P < 0.001$) and between fiber lumen

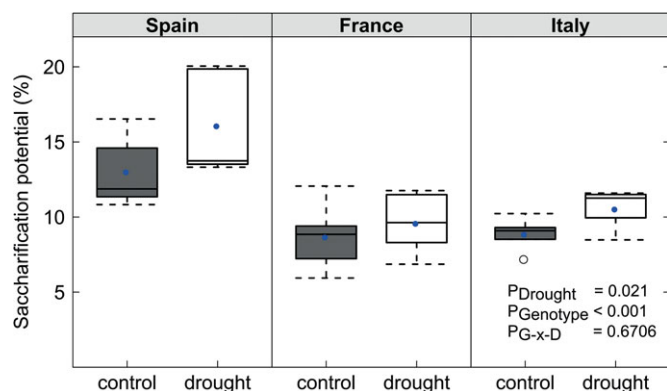


Figure 1. Saccharification potential of well-watered and drought-treated *Populus nigra* genotypes originating from contrasting habitats in Spain, Italy and France. Saccharification potential was expressed as glucose yield per dry cell wall residue ($n = 5$ biological replicates). In the boxes, circles denote the mean and horizontal lines the median. P values for the drought and genotype main effect, and for the genotype–drought interaction effect ($P_{G \times D}$) are given.

and fraction of cell wall area ($P < 0.001$, Figure 2), indicating that decreases in vessel lumen are strongly connected with increases in vessel frequencies, whereas decreases in fiber lumen apparently increased the fraction of cell wall area.

The transcriptome of the developing xylem varied substantially among genotypes and was moderately affected by drought

In order to identify genes and clusters of co-expressed genes underlying genotype or drought-induced variation in saccharification potential and wood traits, whole-transcriptome gene expression profiling of the developing xylem was done by RNA-sequencing. After trimming to remove adaptor contamination and low-quality reads we obtained 23.1 ± 0.95 M reads (mean \pm SE) per sample. After alignment to the *P. trichocarpa* reference and removal of genes with low expression levels, a count table was obtained for 23,393 unique gene models.

Principal component analysis of the count data revealed a prominent separation of genotypes along both of the first two principal components, which together explained 60% of the variance, while clustering according to treatment was less apparent (Figure S2 available as Supplementary Data at *Tree Physiology Online*). Testing for differential expression confirmed strong genotype-related effects because a total of 15,657 GDEGs (67% of the unique gene models tested) were found. Pairwise genotype contrasts between the genotypes showed that $\sim 20\%$ of the GDEGs differed among all genotypes (Figure S3 available as Supplementary Data at *Tree Physiology Online*); the remaining GDEGs showed similar distributions between pairwise comparisons, suggesting similar divergence of the molecular regulation among the three genotypes (Figure S3 available as Supplementary Data at *Tree Physiology Online*).

A drought main effect was detected for only 347 DDEGs, of which 219 were down-, and 128 were up-regulated in response to low water availability. A significant genotype–drought interaction effect was not evident for any of the tested genes.

Gene co-expression modules showed distinct correlations with wood anatomical and chemical traits

In order to relate the sets of GDEGs and DDEGs to genotype-related or drought-induced variation in wood traits, we first reconstructed transcriptional networks of GDEGs and DDEGs by WGCNA and then correlated modules of co-expressed genes with wood traits (Figure 3).

The GDEGs clustered in 18 co-expression modules (DX-G1 to DX-G18; Figure 3A). Wood traits that showed no genotype-related variation (number of cambial cell layers, thickness of fiber and vessel walls, and the fraction of cell wall area; Table 1) showed no or weak correlations with the co-expression modules of GDEGs (Figure 3A). Furthermore, we identified three modules DX-G1, DX-G12 and DX-G7 without any significant correlation with wood traits (Figure 3A). The wood anatomical traits

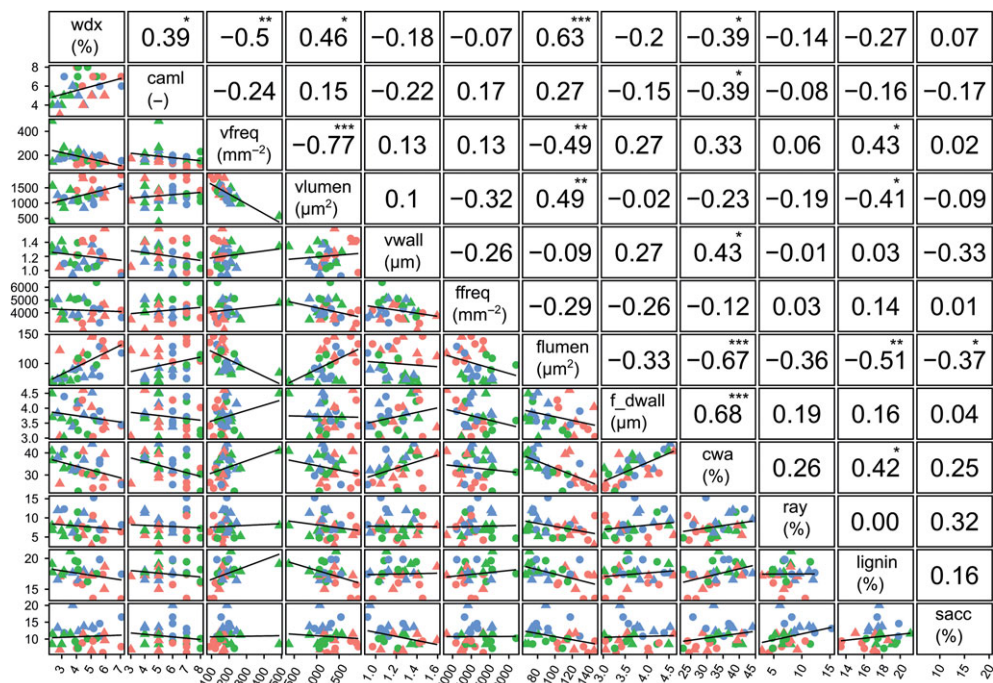


Figure 2. Correlation of wood traits of *Populus nigra* genotypes originating from contrasting habitats in Spain (blue symbols), France (red symbols) and Italy (green symbols) exposed to a control (circles) or drought (triangles) treatment for 5 weeks. Numbers above the diagonal denote Pearson's correlation coefficient r , asterisks denote significance of correlation (*** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$). Lignin, lignin content in dry cell wall; vfreq, vessel frequency; vlumen, average vessel lumen area; vwall, vessel wall thickness; ffreq, fiber frequency; flumen, fiber lumen area; f_dwall, fiber double wall thickness; cwa, percentage of cell wall area of vessels and fibers; wdx, relative radial width of the developing xylem; ray, percentage of ray area; sacc, saccharification potential expressed as glucose yield per dry cell wall residue; caml, number of cambium cell layers. All trait data had the same units as in Table 1.

showed significant positive or negative correlations with DX-G2, DX-G3, DX-G6, DX-G8, DX-G9, DX-G11 and DX-G18 with frequent overlaps (Figure 3A). For example, DX-G3 showed positive correlations with fiber and vessel frequencies and negative correlations with fiber and vessel lumina and the relative width of the developing xylem.

The modules of the GDEGs that were correlated with the saccharification potential (positive: DX-G13, DX-G14, DX-G15, DX-G16, DX-G17; negative: DX-G4, DX-G5, DX-G10) showed in most cases no overlap with wood traits (Figure 3A). Exceptions were DX-G13, which was inversely correlated with vessel lumen and DX-G10, which was inversely correlated with fiber lumen.

To obtain information on the functions that were correlated with the chemical or anatomical traits, the modules were subjected to GO term analyses. DX-G13 and DX-G10, the modules inversely correlated with vessel and fiber lumina, showed significant enrichment of GO terms related to 'catechol-containing compound biosynthesis' and 'cinnamic acid metabolism' (DX-G13), and in 'small molecule catabolic process', 'coumarin biosynthetic process', 'flavonoid biosynthetic process' and 'cell wall biogenesis' (DX-G10) (Table S2 available as Supplementary Data at *Tree Physiology* Online).

The modules positively correlated with saccharification potential showed significant enrichments in GO terms for

'flavonoid biosynthesis' and 'anthocyanin-containing compound biosynthesis' (DX-G14), 'cellular protein modification processes', 'ethylene-related signaling' (DX-G16) and 'regulation of cell morphogenesis and growth' (DX-G17; Table S2 available as Supplementary Data at *Tree Physiology* Online). Modules DX-G10, DX-G4 and DX-G5 were significantly negatively correlated with saccharification potential (see Table S3 available as Supplementary Data at *Tree Physiology* Online for GDEGs assigned to these modules) and were significantly enriched in GO terms for 'carbohydrate metabolic process' (DX-G4, DX-G5), 'cellular metabolic compound salvage' and 'thioester metabolic process' (DX-G5).

The drought-induced DEGs clustered in three co-expressions modules (DX-D1, DX-D2, DX-D3; Figure 3B). Eigengenes of these modules were correlated with those wood anatomical traits that showed significant differences between control and drought treated plants (Figure 3B, Table 1), and with saccharification potential (Figure 3B, Table 1). Module DX-D1 was significantly positively correlated with the number of cambial cell layers ($P = 0.017$) and significantly enriched in GO terms related to 'translation' and 'cellular macromolecule biosynthetic process' (Table S2 available as Supplementary Data at *Tree Physiology* Online). Module DX-D2 was also positively correlated with the number of cambial cell layers ($P = 0.019$) but in addition

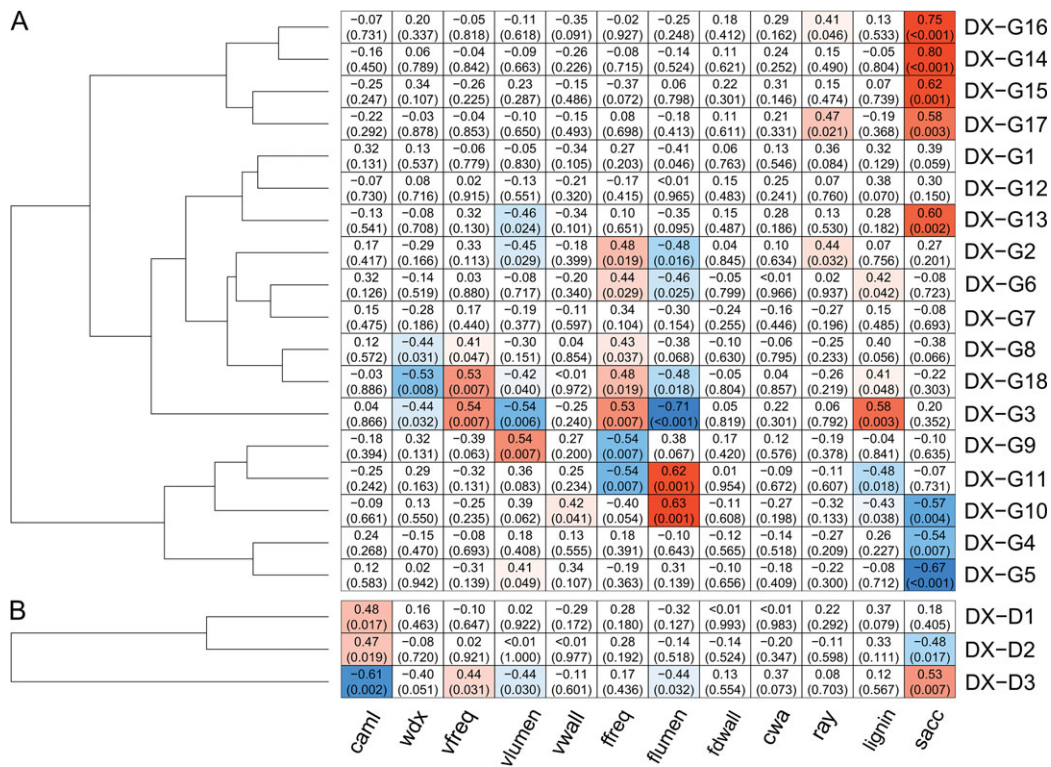


Figure 3. Heatmap representing the correlation of eigengenes of co-expression modules with wood traits for genes showing significant genotype-related variation in transcript abundance (DX-G1 to DX-G18) (A) or drought-responsive differentially expressed genes (DX-D1 to DX-D3) (B). Eigengenes of the modules correspond to the first principal component. Numbers represent Pearson's correlation coefficient r and the P value for the correlation (in brackets). For abbreviations of traits see Figure 2.

significantly negatively correlated with saccharification potential ($P = 0.017$, Figure 3B). Module DX-D3 was significantly negatively correlated with the number of cambial cell layers ($P = 0.002$), fiber lumen area ($P = 0.032$) and vessel lumen area ($P = 0.03$), but strongly positively correlated with saccharification potential ($P = 0.007$, Figure 3B). The GO terms associated with this module pointed to membrane-related processes ('membrane-bounded organelle' and 'intracellular membrane-bounded organelle').

To identify putative key genes involved in the control of saccharification potential, we ranked all GDEGs and DDEGs assigned to modules significantly correlated with saccharification potential (Figure 3) by the significance of correlation of transcript abundance with this wood property. The most strongly correlated GDEGs were protein kinase proteins, MAPKKK proteins, cytochrome P450 proteins, a heteroglucan glucosidase and others (Table S3 available as Supplementary Data at *Tree Physiology* Online, Figure 4A). The DDEGs most strongly correlated with saccharification potential included nucleotide-sugar transporter family proteins, a phosphomannomutase, a methyltransferase and others (Table S4 available as Supplementary Data at *Tree Physiology* Online, Figure 4B).

Metabolic processes in the developing xylem underlying variation in saccharification potential

Since variation in saccharification potential was largely independent from wood anatomical traits (Figure 2), we sought to get additional insights into the metabolic control of saccharification potential by mapping all GDEGs that were assigned to any of the saccharification-correlated modules and whose transcript abundance was significantly correlated with the saccharification potential onto MapMan functional categories ('bins').

An over-representation of GDEGs with strong positive or negative correlations to saccharification potential was detected in five bins, namely 'cell/cell vesicle transport', 'cell wall', 'misc/cytochrome P450' and 'protein targeting secretory pathway' (Table 2). A closer inspection of the bin 'cell wall' revealed that transcript abundances of GDEGs involved in 'cell wall modification' (bin 10.7), 'hemicellulose synthesis' (bin 10.3), 'pectin synthesis' (bin 10.4), 'cellulose synthesis' (bin 10.2) and 'cell wall degradation' (bins 10.6.2 and 10.6.3) were mostly negatively correlated with saccharification potential (Figure 5A, Table S5 available as Supplementary Data at *Tree Physiology* Online). In contrast, bins 'pectin esterases' (bin 10.8) and 'cell wall/precursor synthesis' (bin 10.1) included specific processes that were positively correlated with saccharification

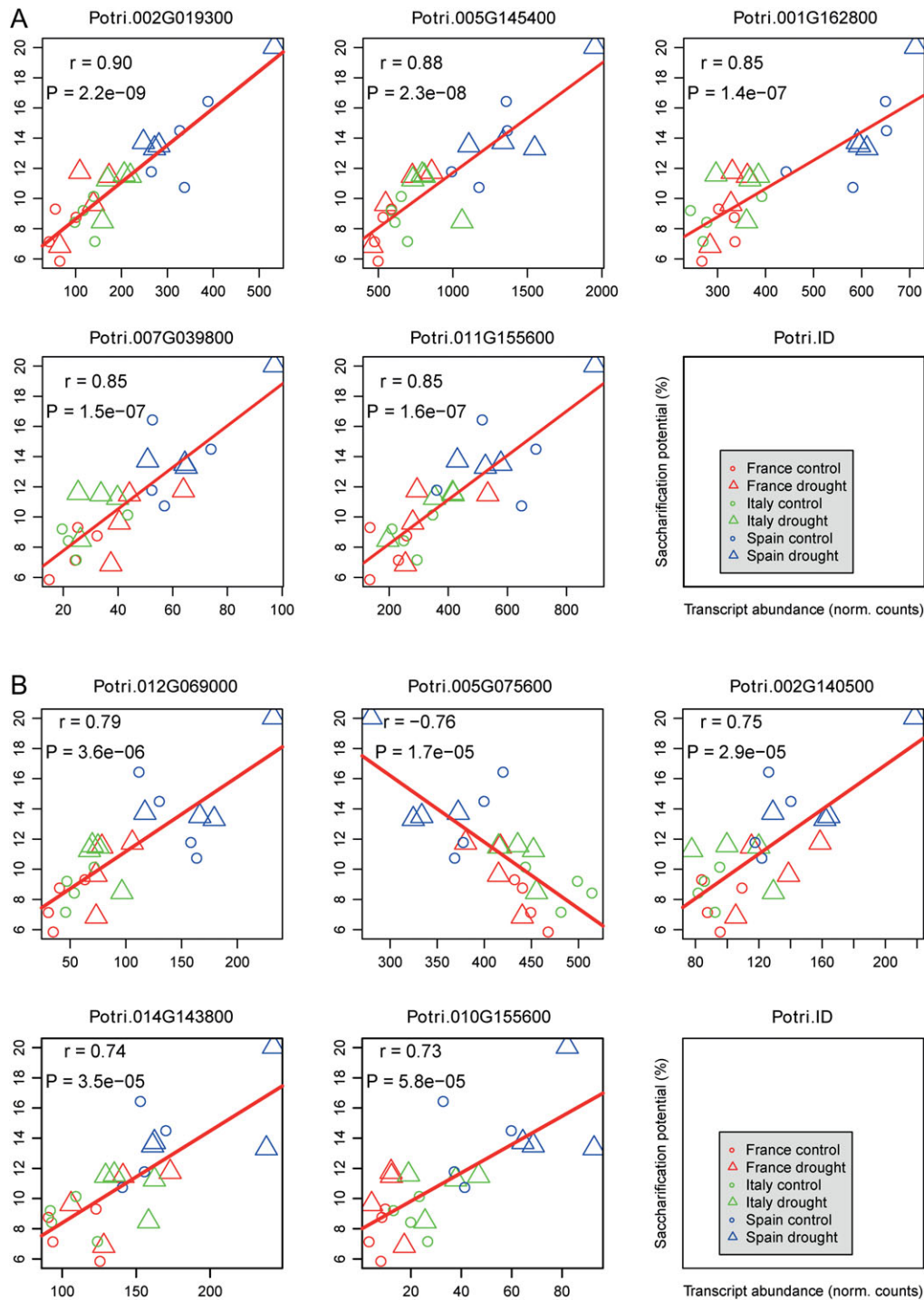


Figure 4. Plots of saccharification potential against transcript abundance of genotype-related differentially expressed genes (GDEGs, A) and drought-induced differentially expressed genes (DDEGs, B). The top five most significantly correlated GDEGs and DDEGs with saccharification potential are shown. All genes show a significant genotype or drought main effect and are assigned to saccharification related co-expression modules. Pearson's correlation coefficient r and the P value for the correlation are given.

potential. The bin 'cell wall/precursor synthesis' (Figure 5B) contained transcripts encoding UDP-D-glucose/UDP-D-galactose epimerases (bin 10.1.2), UDP-glucuronic acid decarboxylases (bin 10.1.5), and enzymes involved in GDP-D-mannose (bins 10.1.20, 10.1.21, 10.1.1) and UDP-rhamnose biosynthesis (bin 10.1.11)

that were negatively correlated with saccharification potential. In contrast, transcripts coding for enzymes involved in the biosynthesis of UDP-L-arabinose (bin 10.1.9, UDP-arabinose 4-epimerase), UDP-D-galacturonic acid (bin 10.1.6, UDP-D-glucuronate 4-epimerase 2) and UDP-L-rhamnose (bin 10.1.10,

Table 2. MapMan bins enriched in genes with transcript abundance highly correlated to variation in saccharification potential in three genotypes of *Populus nigra*.

Bin ID	Bin name	Number of genes mapped	P value
31	Cell	194	2.86E-08
31.4	Cell/vesicle transport	52	8.60E-07
10	cell wall	89	2.50E-03
26.10	Misc/cytochrome P450	25	2.83E-02
29.3.4	Protein targeting/secretory pathway	52	3.48E-02

P values as calculated by MapMan's in-built Wilcoxon rank sum test and adjusted for multiple testing by Benjamini-Yekutieli correction. Mapping was done with all genes which (i) differed significantly in expression among genotypes, (ii) were assigned to gene co-expression modules significantly correlated with saccharification potential and (iii) were significantly correlated to saccharification on the level of their individual transcript abundance.

UDP-L-rhamnose synthase) were positively correlated with saccharification potential (Figure 5B; Table S5 available as Supplementary Data at *Tree Physiology* Online).

In order to identify regulatory genes commonly co-expressed with saccharification-related GDEGs assigned to the MapMan bin 'cell wall' we used the latter genes as baits in an in silico co-expression analysis. By this approach we identified 2510 genes significantly co-expressed with the bait genes, and 827 of these co-expressed genes were among our initial set of saccharification related GDEGs. The overlapping genes included 31 unique transcription factors, including MYB domain transcription factors, WRKY transcription factors, homeobox protein transcription factors and others (Table 3).

Discussion

Drought-induced variations in wood anatomical traits are independent from genotype

Drought significantly reduced, but did not abolish diameter increment in all genotypes, along with a reduction in the number of cambial cell layers (Table 1), indicating a reduced cambial activity under these conditions. These findings agree with previous studies (Arend and Fromm 2007, Bogeat-Triboulot et al. 2007) and demonstrate that the moderate, gradually increasing drought, applied in a highly controlled manner, was effective to induce acclimation processes. Such moderate yet significant treatment effects mimicking ecologically relevant stress conditions are suited to study the molecular control of drought-induced and genotypic variation in wood traits.

Along with a reduction in cambial activity, a significant increase in the fraction of cell wall area of vessels, fibers and rays under drought was noted, while vessel and fiber wall thickness were not affected by drought in any of the genotypes (Table 1). Interestingly, the three *P. nigra* genotypes studied

here showed similar drought-induced changes in wood anatomy, although the wood structures clearly differed among the genotypes. Constitutive differences were also apparent among the transcriptomes because the majority, i.e., 67% of all tested genes, showed genotype-related expressional differences in the developing xylem. This observation underlines that substantial genotypic variations in transcript abundance exist not only in leaves and roots (leaves: Wilkins et al. 2009, roots and leaves: Cohen et al. 2010, leaves: Hamanishi et al. 2010, DeWoody et al. 2015) but also in wood-forming tissues (this study). Under the current experimental conditions, which allowed gradual acclimation to drought, surprisingly no genotype × drought interaction was detected for any gene. Thus, differences in wood anatomy are the result of constitutive differences among the transcriptomes, while flexible adjustment of wood anatomy to variation in water availability is apparently mediated by similar molecular programs.

Drought-induced and genotype-related variations in wood anatomical traits and lignin content are not related to variation in saccharification potential

In contrast to our initial hypothesis that drought will cause an increase in lignification (Moore et al. 2008, Le Gall et al. 2015), the lignin content was not affected under drought. We had further hypothesized that along with an increased lignification, drought will cause a reduction in saccharification potential. Our data do not support these hypotheses, since we observed a significant increase in saccharification potential under drought across all genotypes (Figure 1) and no correlation with lignin or with wood anatomy. We are not aware of any previous studies on drought effects on saccharification potential in woody dicots, but van der Weijde et al. (2016) reported an increase in saccharification potential of *Miscanthus* under drought. These observations suggest that drought may cause alterations in cell wall chemistry apart from lignin that facilitate the enzymatic release of glucose.

Important bioenergy crops such as *Miscanthus* (Souza et al. 2015), *Eucalyptus* (Klash et al. 2010, Elissetche et al. 2011), and *Populus* (Wegrzyn et al. 2010, Studer et al. 2011, Zhou et al. 2011, Guerra et al. 2013, Porth et al. 2013, Muchero et al. 2015, Fahrenkrog et al. 2016) exhibit substantial natural genetic variation in lignin content. Likewise, saccharification potential of bioenergy crops showed genetic variation in earlier studies (Studer et al. 2011, Souza et al. 2015, van der Weijde et al. 2016). In agreement with these findings, *P. nigra* also showed genotypic variations in lignin content and in saccharification potential (Table 1, Figure 1). Our hypothesis that adaptation of poplars to low water availability in their native habitat correlates with high lignin content and low saccharification potential has to be refuted: saccharification potential was unexpectedly highest in genotype Spain, originating from the driest habitat, whereas the lignin content of the Spain genotype was not different from that of the Italy genotype, which originated

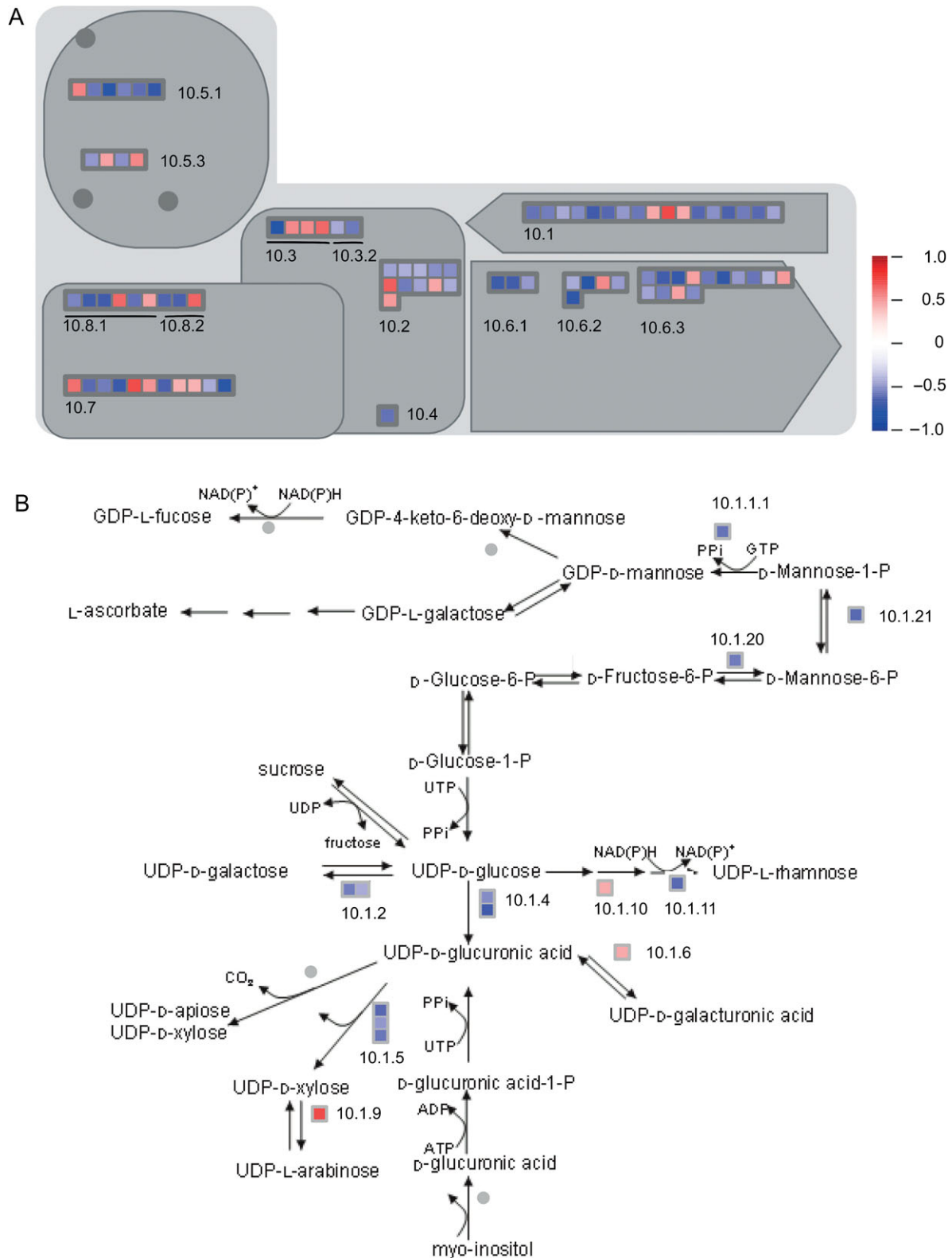


Figure 5. Mapping of saccharification-related genes to Map Man functional categories (bins) related to cell wall metabolism (A, overview; B, details of bin 10.1, 'precursor synthesis'). Numbers in (A) refer to cell wall-related bins as follows: 10.1, 'precursor synthesis'; 10.2, 'cellulose synthesis'; 10.3, 'hemicellulose synthesis'; 10.4, 'pectin synthesis'; 10.5, 'cell wall proteins'; 10.6, 'degradation'; 10.7, 'modification', 10.8, 'pectin esterases' and in (B) to the bin 'cell wall/precursor synthesis' with sub-bins explained in Table S5 (available as Supplementary Data at *Tree Physiology Online*). Color code represents Pearson's correlation coefficient of transcript abundance with saccharification potential.

Table 3. Saccharification-related bait genes present in the functional category 'cell wall' and prey transcription factors.

Bait genes					Co-expressed <i>Arabidopsis</i> genes						
Ath locus	Ath hit description	Ptr locus	r_{sacc}	$P(r_{\text{sacc}})$	Locus	Description	r^2_{coexp}	$P(r^2_{\text{coexp}})$	Ptr locus	r_{sacc}	$P(r_{\text{sacc}})$
AT2G42800	Receptor like protein 29	Potri.008G211800	-0.48	1.63E-02	AT1G23380	KNAT6	0.87	9.88E-52	Potri.008G188700	0.49	1.55E-02
AT2G23130	Arabinogalactan protein 17	Potri.007G051600	0.56	4.63E-03	AT1G29160	Dof-type zinc finger domain-containing protein	0.81	1.54E-43	Potri.011G065900	-0.50	1.38E-02
AT2G42800	Receptor like protein 29	Potri.008G211800	-0.48	1.63E-02	AT1G53160	SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 4	0.86	3.35E-50	Potri.007G138800	-0.48	1.81E-02
AT2G42800	Receptor like protein 29	Potri.008G211800	-0.48	1.63E-02	AT1G75430	BEL1-LIKE HOMEODOMAIN 11 (BLH11)	0.83	9.01E-46	Potri.002G030900	0.60	2.03E-03
AT2G36870	Xyloglucan endotransglucosylase/hydrolase 32	Potri.009G006600	0.51	1.08E-02	AT1G77920	bZIP family transcription factor	0.85	7.85E-50	Potri.002G090700	0.63	9.38E-04
AT2G40610	Expansin A8	Potri.013G154700	0.42	4.00E-02	AT2G18300	Basic helix-loop-helix (bHLH) family protein	0.83	7.23E-46	Potri.007G023600	0.44	3.22E-02
AT1G02810	Plant invertase/pectin methylesterase inhibitor superfamily	Potri.014G127000	-0.54	6.82E-03	AT2G34830	WRKY DNA-binding protein 35 (WRKY35)	0.86	5.49E-51	Potri.002G195300	0.65	6.08E-04
AT2G42800	Receptor like protein 29	Potri.008G211800	-0.48	1.63E-02	AT2G42280	bHLH family protein	0.83	1.86E-45	Potri.006G057600	0.57	3.41E-03
AT2G47930	Arabinogalactan protein 26	Potri.002G207500	-0.57	3.45E-03	AT2G47260	WRKY23	0.81	7.34E-43	Potri.014G118200	0.59	2.28E-03
AT2G42800	Receptor like protein 29	Potri.008G211800	-0.48	1.63E-02	AT3G13960	GROWTH-REGULATING FACTOR 5 (AtGRF5)	0.82	3.80E-45	Potri.019G042300	-0.73	4.35E-05
AT2G42800	Receptor like protein 29	Potri.008G211800	-0.48	1.63E-02	AT3G15270	SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 5	0.88	2.99E-54	Potri.001G398200	0.64	7.32E-04
AT2G42800	Receptor like protein 29	Potri.008G211800	-0.48	1.63E-02	AT3G19184	DNA binding	0.88	1.24E-53	Potri.005G137600	-0.64	7.97E-04
AT2G36870	Xyloglucan endotransglucosylase/hydrolase 32	Potri.009G006600	0.51	1.08E-02	AT3G28920	ARABIDOPSIS THALIANA HOMEBOX PROTEIN 34	0.81	4.86E-43	Potri.015G032700	0.62	1.17E-03
AT1G02810	Plant invertase/pectin methylesterase inhibitor superfamily	Potri.014G127000	-0.54	6.82E-03	AT3G49760	<i>Arabidopsis thaliana</i> basic leucine-zipper 5 (AtbZIP5)	0.85	1.51E-48	Potri.007G006900	0.52	9.76E-03
AT2G42800	Receptor like protein 29	Potri.008G211800	-0.48	1.63E-02	AT4G13640	Unfertilized embryo sac 16 (UNE16)	0.80	2.56E-42	Potri.001G314800	0.50	1.25E-02
AT2G42800	Receptor like protein 29	Potri.008G211800	-0.48	1.63E-02	AT4G37740	GROWTH REGULATING FACTOR 2 (AtGRF2)	0.81	3.85E-43	Potri.003G100800	-0.69	1.92E-04
AT2G42800	Receptor like protein 29	Potri.008G211800	-0.48	1.63E-02	AT4G37750	AINTEGUMENTA (ANT)	0.93	3.75E-69	Potri.014G012200	0.47	2.06E-02
AT1G32170	Xyloglucan endotransglucosylase/hydrolase 30	Potri.003G097300	-0.61	1.51E-03	AT5G02030	REPLUMLESS (RPL)	0.86	5.29E-51	Potri.010G197300	0.50	1.34E-02
AT1G11580	Methylesterase PCR A	Potri.011G135000	-0.68	2.91E-04	AT5G02030	REPLUMLESS (RPL)	0.90	1.69E-59	Potri.010G197300	0.50	1.34E-02
AT2G42800	Receptor like protein 29	Potri.008G211800	-0.48	1.63E-02	AT5G03790	HB51	0.84	1.17E-46	Potri.006G117700	-0.54	5.94E-03
AT2G28760	UDP-XYL synthase 6	Potri.010G207200	-0.62	1.36E-03	AT5G05790	myb family transcription factor	0.86	6.09E-50	Potri.010G193000	-0.57	3.63E-03
AT1G11580	Methylesterase PCR A	Potri.011G135000	-0.68	2.91E-04	AT5G05790	myb family transcription factor	0.84	1.09E-47	Potri.010G193000	-0.57	3.63E-03

(Continued)

Table 3. (Continued)

Bait genes					Co-expressed <i>Arabidopsis</i> genes						
Ath locus	Ath hit description	Ptr locus	r_{sacc}	$P(r_{\text{sacc}})$	Locus	Description	r^2_{coexp}	$P(r^2_{\text{coexp}})$	Ptr locus	r_{sacc}	$P(r_{\text{sacc}})$
AT1G11580	Methylesterase PCR A	Potri.011G135000	-0.68	2.91E-04	AT5G06800	myb family transcription factor	0.83	3.00E-45	Potri.006G191000	-0.45	2.77E-02
AT1G02810	Plant invertase/pectin methylesterase inhibitor superfamily	Potri.014G127000	-0.54	6.82E-03	AT5G10280	MYB DOMAIN PROTEIN 92 (ATMYB92)	0.83	5.55E-46	Potri.007G093900	-0.43	3.41E-02
AT2G23130	Arabinogalactan protein 17	Potri.007G051600	0.56	4.63E-03	AT5G12330	LATERAL ROOT PRIMORDIUM 1 (LRP1)	0.81	2.59E-43	Potri.003G196100	0.65	5.87E-04
AT3G22800	Leucine-rich repeat (LRR) family protein	Potri.010G083000	-0.50	1.31E-02	AT5G14750	MYB DOMAIN PROTEIN 66 (ATMYB66)	0.81	1.11E-42	Potri.003G064600	0.53	7.37E-03
AT1G02810	Plant invertase/pectin methylesterase inhibitor superfamily	Potri.014G127000	-0.54	6.82E-03	AT5G14750	MYB DOMAIN PROTEIN 66 (ATMYB66)	0.91	1.44E-60	Potri.003G064600	0.53	7.37E-03
AT2G36870	Xyloglucan endotransglucosylase/hydrolase 32	Potri.009G006600	0.51	1.08E-02	AT5G15210	ARABIDOPSIS THALIANA HOMEBOX PROTEIN 30	0.82	8.67E-44	Potri.004G126500	-0.48	1.86E-02
AT2G42800	Receptor like protein 29	Potri.008G211800	-0.48	1.63E-02	AT5G37020	AUXIN RESPONSE FACTOR 8	0.94	4.72E-70	Potri.001G358500	-0.53	8.37E-03
AT1G11580	Methylesterase PCR A	Potri.011G135000	-0.68	2.91E-04	AT5G52830	WRKY27	0.81	1.06E-42	Potri.017G149000	0.51	1.16E-02
AT1G32170	Xyloglucan endotransglucosylase/hydrolase 30	Potri.003G097300	-0.61	1.51E-03	AT5G57620	myb domain protein 36 (MYB36)	0.83	5.74E-46	Potri.006G123400	0.46	2.47E-02
AT1G11580	Methylesterase PCR A	Potri.011G135000	-0.68	2.91E-04	AT5G57620	myb domain protein 36 (MYB36)	0.80	2.41E-42	Potri.006G123400	0.46	2.47E-02
AT2G42800	Receptor like protein 29	Potri.008G211800	-0.48	1.63E-02	AT5G60910	Agamous-like 8 (AGL8)	0.82	8.32E-45	Potri.003G170200	0.55	5.00E-03
AT2G28760	UDP-XYL synthase 6	Potri.010G207200	-0.62	1.36E-03	AT5G65210	TGA1	0.84	1.01E-46	Potri.007G085700	0.70	1.43E-04
AT1G11580	Methylesterase PCR A	Potri.011G135000	-0.68	2.91E-04	AT5G65210	TGA1	0.91	9.93E-61	Potri.007G085700	0.70	1.43E-04
AT2G42800	Receptor like protein 29	Potri.008G211800	-0.48	1.63E-02	AT5G66770	Scarecrow transcription factor family protein	0.83	5.23E-46	Potri.005G123800	0.76	1.87E-05

Co-expression of *Arabidopsis* prey genes with bait genes was analyzed across multiple microarray experiments using CressExpress (<http://cressexpress.org/>). Ath Locus: *Arabidopsis thaliana* gene locus; Ptr Locus: gene locus IDs in the Phytozome *Populus trichocarpa* v3.0 database; r_{sacc} : Pearson's coefficient for the correlation of *Populus nigra* best hits' gene transcript abundance with saccharification potential; $P(r_{\text{sacc}})$: P -values for the respective correlations. r^2_{coexp} and $P(r^2_{\text{coexp}})$: squared Pearson's correlation coefficient and P values for the correlation of the transcript abundances of a bait with co-expressed *Arabidopsis* gene.

from the wettest habitat (Table 1, Figure 1). Previous studies also reported inconsistent results concerning the relationship between lignin content and saccharification potential. Some studies with transgenic poplars with low lignin content reported negative correlations between lignin and saccharification potential (Min et al. 2012, Acker et al. 2014), while Voelker et al. (2010) found unaffected saccharification potential when lignin decrease was modest, and decreasing saccharification in transgenic poplar lines with strongly reduced lignin content. Studer et al. (2011) did not find a general negative relationship between natural variation in lignin content of *P. trichocarpa* and sugar release, but reported that the effect of lignin on saccharification depends on the lignin composition and pretreatment methods. In *Miscanthus* natural genetic variation in lignin content was not consistently correlated with saccharification potential (Souza et al. 2015). In a study with maize, QTLs for lignin content were different from those detected for saccharification (Penning et al. 2014). Altogether, these studies suggest that lignin is not a good predictor for the saccharification potential.

Transcriptional regulation of genes involved in cell wall polysaccharide biosynthesis and modification is related to genotypic variation in saccharification potential

To uncover the molecular basis for differences in saccharification potential we re-constructed transcriptional networks and found that co-expression modules correlated with this trait were enriched in distinct sets of GO terms (Table S2 available as Supplementary Data at *Tree Physiology* Online). This finding underpins the assumption that modules summarize functionally related genes. The functional categories pointed to an involvement of cell wall metabolism, especially cell wall polysaccharides (bin 10), and vesicle-associated secretory processes (bins 31.4 and 29.3.4) in the control of saccharification potential. In agreement with lacking correlation between lignin content and saccharification potential, categories related to lignin biosynthesis were not enriched, suggesting that variation in expression of genes involved in cell wall polysaccharide biosynthesis is more important for saccharification potential than moderate variations in lignin. However, among the top saccharification-correlated genes, none of the biosynthetic genes for cell wall polysaccharides was detected. Instead three putative kinases (Potri.002G019300, Potri.007G039800, Potri.010G155600) and one gene involved in brassinosteroid metabolism (Potri.011G155600) were present among the top 10 correlated genes (Figure 4), suggesting that signaling and hormone regulation are important check points for the molecular control of saccharification potential.

Surprisingly, most GDEGs assigned to cellulose biosynthesis (bins 10.2, 10.2.1, 10.2.2) were negatively correlated with saccharification potential (Figure 5A, Table S5 available as Supplementary Data at *Tree Physiology* Online). Saccharification potential depends not only on cellulose content, but also on the level of cellulose aggregation and thus accessibility for enzymatic

degradation (Nookaraju et al. 2013). Primary cell walls contain cellulose microfibrils with low aggregation, whereas during the synthesis of secondary cell walls, highly aggregated microfibrils are formed (Li et al. 2016). Synthesis of cellulose in primary and secondary cell walls involves distinct sets of genes (Li et al. 2016). It is thus possible that the GDEGs that were negatively correlated with saccharification potential (Table S5 available as Supplementary Data at *Tree Physiology* Online) represent poplar cellulose synthase genes involved in the synthesis of highly aggregated cellulose microfibrils.

Although pectins comprise only a small fraction of secondary cell walls (Harholt et al. 2010) there is accumulating evidence that they play a role in determining saccharification potential, by mediating cell adhesion, masking cellulose and hemicelluloses and thus restricting their accessibility by enzymes (Xiao and Anderson 2013, Tavares et al. 2015). The effect of pectins on saccharification is dependent on the degree of pectin acetylation and methyl-esterification (Xiao and Anderson 2013). Once secreted into the apoplast, pectins can be de-methylesterified by pectin methylesterases, enabling cross-linking and possibly increasing cell wall stiffness (Xiao and Anderson 2013). A reduction of de-methylesterified pectins was found to increase saccharification potential in *Arabidopsis*, wheat and tobacco (Lionetti et al. 2010). In agreement with these findings, putative pectin methylesterases (bins 10.8.1 and 10.8.2) were mostly negatively correlated with saccharification potential (Figure 5, Table S5 available as Supplementary Data at *Tree Physiology* Online).

The major compound of hemicellulose and second most abundant polysaccharide in woody dicots is glucuronoxylan (Awano et al. 1998, Lee et al. 2009). Here the transcript abundances of two genes related to glucuronoxylan biosynthesis (Potri.002G132900 and Potri.016G086400, of bin 10.3.2) were negatively correlated with saccharification potential (Figure 5A, Table S5 available as Supplementary Data at *Tree Physiology* Online). There is experimental evidence that the *P. × canescens* orthologs of Potri.002G132900 and Potri.016G086400 are functional homologs of *Arabidopsis* PARVUS (Lee et al. 2009) and IRX9 (Zhou et al. 2007, Lee et al. 2011), respectively. PARVUS (At1g19300) and IRX9 (At2g37090) are required for the synthesis of the tetrasaccharide reducing end sequence of glucuronoxylans, and the xylosyl backbone of glucuronoxylans, respectively (Brown et al. 2005, 2007, Lee et al. 2007a, 2007b, Peña et al. 2007). Lee et al. (2011) found that RNAi down-regulation of the *P. × canescens* ortholog of Potri.016G086400 resulted in an increased glucose release from wood. Remarkably, UDP-glucuronic acid decarboxylase (UXS) transcript abundances were negatively correlated with saccharification potential (GDEGs in bin 10.1.5, Figure 5B, Table S5 available as Supplementary Data at *Tree Physiology* Online). UXS enzymes catalyze the irreversible synthesis of UDP-xylose from UDP-glucuronic acid (Kuang et al. 2016), and are key enzymes for nucleotide sugar partitioning for cell wall

polysaccharide biosynthesis (Du et al. 2013). In transgenic tobacco, down-regulation of UXS genes led to a significant increase in saccharification efficiency (Cook et al. 2012). Taken together, these findings illustrate the power of our experimental and analytical approach to identify candidate genes related to the regulation of wood traits.

Putative transcriptional regulators of saccharification-related genes involved in cell wall metabolism

To identify putative transcriptional regulators of saccharification-related genes involved in cell wall metabolism, we used the identified cell wall genes as baits in an *in silico* co-expression analysis. Among the genes with a significant correlation with saccharification, 31 potential transcription factors were identified that showed significant co-expression with cell wall biosynthesis genes in *Arabidopsis* and were also present among the poplar genes, significantly correlated with saccharification (Table 3). Among these genes, MYB domain, WRKY and homeobox domain transcription factors that belong to families known to be involved in wood formation were retrieved (Hussey et al. 2013, Yang and Wang 2016). For example, two putative pectin esterases (Potri.014G127000, Potri.011G135000) were co-expressed with WRKY, MYB and bZIP transcription factors. Interestingly, a gene encoding a putative UXS protein (Potri.010G207200), which appears to be important for the saccharification potential as discussed above, was co-expressed with a MYB family (Potri.010G193000) and TGA1 (Potri.007G085700) transcription factor. MYB transcription factors are known to play a role in the regulation of xylan biosynthesis (Rennie and Scheller 2014), but to our knowledge, no published results are available for the specific MYB family members highlighted by our study.

In addition, the gene set of co-expressed transcription factors contained novel candidate regulators such as TGACG SEQUENCE-SPECIFIC BINDING PROTEIN 1 (TGA1), SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 4 (SPL4), AGAMOUS-LIKE 8 (AGL8), SCARECROW, AUXIN RESPONSE FACTOR 8 (ARF8), ANTEGUMENTA (ANT), etc. These transcription factors play roles in redox regulation (TGA) and development. Our approach suggests that developmental changes brought about by drought or genetic differentiation among the genotypes also control cell wall composition that affects saccharification. Therefore, the present correlative results provide an excellent platform for future targeted approaches to improve the saccharification potential of poplar.

Conclusions

The efficient conversion of lignocellulosic biomass to biofuels varies with wood composition, which itself is subject to natural genetic variation and is dependent on environmental conditions, such as water availability. Previous research on wood traits that may affect saccharification mainly focused on lignin, revealing opportunities for improving biomass quality by reducing lignin

content. However, there is also growing awareness that this approach has limitations (Voelker et al. 2010, Tavares et al. 2015). By studying variation in wood traits and saccharification potential in three genotypes of *P. nigra* exposed to a moderate drought treatment, we show that glucose release was not impaired, but moderately improved, by gradually decreasing water availability. Interestingly, the saccharification potential was not related to lignin content or expression of genes related to lignin biosynthesis. Instead, our research identified transcriptional regulation of biosynthesis of hemicelluloses and modification of pectins as potential targets for improving wood quality for the production of biofuels. Further experiments are needed in which the expression of the identified candidate genes is modified by overexpression or suppression to test the causal relationship of the identified genes and pathways with saccharification potential.

Supplementary Data

Supplementary Data for this article are available at *Tree Physiology* Online.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author contributions

H.W., S.P., H.K.S., S.K.S., M.J.P., S.S., D.J., C.D., D.C., I.H., D.L.T., O.B., J.F., M.-B.B.-T., G.T. and A.P. designed the experiment. H. W., S.P., H.K.S., D.J., C.D., D.C., C.B., D.L.T., O.B. and M.-B.B.-T. conducted the greenhouse experiment. M.A. and H.K.S. analyzed saccharification potential, M.Ma. analyzed lignin content and S.P. analyzed wood anatomy. V.V. and F.C. conducted the RNA-sequencing. H.W., D.J., S.K.S., M.J.P. and S.S. conducted bioinformatic and statistical analyses. H.W., M.-B.B.-T., F.v.E., J.J.B.K., J.F., M.Mo., P.R., G.T. and A.P. supervised research. H.W., S.P. and A.P. wrote the manuscript. All authors contributed to and approved the final version of the manuscript.

Funding

This work was conducted in the frame of the WATBIO (Development of improved perennial biomass crops for water-stressed environments), which is a collaborative research project funded from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No. 311929. The research has received funding from the French National Research Agency through the Laboratory of Excellence ARBRE (ANR-12-LABXARBRE-01). S.P. received a Ph.D. scholarship from the program Erasmus Mundus, India4EU11.

Acknowledgments

We thank C. Kettner (Forest Botany and Tree Physiology, Göttingen, Germany) for plant maintenance and plant production and T. Klein (Laboratory for Radio-Isotopes, Göttingen, Germany) for RNA isolation. We acknowledge the providers of the original *P. nigra* genotypes 'France 6J-29' (INRA, Paris, France represented by G. Pilate) and 'Spain RIN2-new' (CITA, Zaragoza, Spain, represented by JV Lacasa Azlor) and C. Bastien (INRA, Orleans, France) for providing the stock cuttings.

References

- Acker RV, Leplé J-C, Aerts D et al. (2014) Improved saccharification and ethanol yield from field-grown transgenic poplar deficient in cinnamoyl-CoA reductase. *Proc Natl Acad Sci USA* 111:845–850.
- Allwright MR, Taylor G (2016) Molecular breeding for improved second generation bioenergy crops. *Trends Plant Sci* 21:43–54.
- Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11:R106.
- Anders S, Pyl PT, Huber W (2015) HTSeq – a python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169.
- Arend M, Fromm J (2007) Seasonal change in the drought response of wood cell development in poplar. *Tree Physiol* 27:985–992.
- Ashburner M, Ball CA, Blake JA et al. (2000) Gene Ontology: tool for the unification of biology. *Nat Genet* 25:25–29.
- Awano T, Takabe K, Fujita M (1998) Localization of glucuronoxylans in Japanese beech visualized by immunogold labelling. *Protoplasma*, 202:213–222.
- Bauer S, Grossmann S, Vingron M, Robinson PN (2008) Ontologizer 2.0 – a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinformatics* 24:1650–1651.
- Beniwal RS, Langenfeld-Heyser R, Polle A (2010) Ectomycorrhiza and hydrogel protect hybrid poplar from water deficit and unravel plastic responses of xylem anatomy. *Environ Exp Bot* 69:189–197.
- Bogeat-Triboulot MB, Brosché M, Renaut J et al. (2007) Gradual soil water depletion results in reversible changes of gene expression, protein profiles, ecophysiology, and growth performance in *Populus euphratica*, a poplar growing in arid regions. *Plant Physiol* 143: 876–892.
- Bourgon R, Gentleman R, Huber W (2010) Independent filtering increases detection power for high-throughput experiments. *Proc Natl Acad Sci USA* 107:9546–9551.
- Brereton NJB, Pitre FE, Hanley SJ, Ray MJ, Karp A, Murphy RJ (2010) QTL mapping of enzymatic saccharification in short rotation coppice willow and its independence from biomass yield. *BioEnergy Res* 3:251–261.
- Brown DM, Zeef LAH, Ellis J, Goodacre R, Turner SR (2005) Identification of novel genes in *Arabidopsis* involved in secondary cell wall formation using expression profiling and reverse genetics. *Plant Cell* 17:2281–2295.
- Brown DM, Goubet F, Wong VW, Goodacre R, Stephens E, Dupree P, Turner SR (2007) Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis. *Plant J* 52: 1154–1168.
- Cao X, Jia J, Zhang C, Li H, Liu T, Jiang X, Polle A, Peng C, Luo Z-B (2014) Anatomical, physiological and transcriptional responses of two contrasting poplar genotypes to drought and re-watering. *Physiol Plant* 151:480–494.
- Capron A, Chang XF, Hall H, Ellis B, Beatson RP, Berleth T (2013) Identification of quantitative trait loci controlling fibre length and lignin content in *Arabidopsis thaliana* stems. *J Exp Bot* 64:185–197.
- Chang S, Puryear J, Cairney J (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Mol Biol Rep* 11: 113–116.
- Cohen D, Bogeat-Triboulot MB, Tisserant E et al. (2010) Comparative transcriptomics of drought responses in *Populus*: a meta-analysis of genome-wide expression profiling in mature leaves and root apices across two genotypes. *BMC Genomics* 11:630.
- Cook CM, Daudi A, Millar DJ, Bindschedler LV, Khan S, Bolwell GP, Devoto A (2012) Transcriptional changes related to secondary wall formation in xylem of transgenic lines of tobacco altered for lignin or xylan content which show improved saccharification. *Phytochemistry* 74:79–89.
- Da Costa RMF, Lee SJ, Allison GG, Hazen SP, Winters A, Bosch M (2014) Genotype, development and tissue-derived variation of cell-wall properties in the lignocellulosic energy crop *Miscanthus*. *Ann Bot* 114:1265–1277.
- DeWoody J, Trewin H, Taylor G (2015) Genetic and morphological differentiation in *Populus nigra* L.: isolation by colonization or isolation by adaptation? *Mol Ecol* 24:2641–2655.
- Du Q, Pan W, Tian J, Li B, Zhang D (2013) The UDP-glucuronate decarboxylase gene family in *Populus*: structure, expression, and association genetics. *PLOS ONE* 8:e60880.
- Eiissethe JP, Valenzuela S, García R, Norambuena M, Iturra C, Rodríguez J, Mendonça RT, Balocchi C (2011) Transcript abundance of enzymes involved in lignin biosynthesis of *Eucalyptus globulus* genotypes with contrasting levels of pulp yield and wood density. *Tree Genet Genomes* 7:697–705.
- Fahrenkrog AM, Neves LG, Resende MF et al. (2016) Genome-wide association study reveals putative regulators of bioenergy traits in *Populus deltoides*. *New Phytol* 213:799–811.
- Fiasconaro ML, Gogorcena Y, Muñoz F, Andueza D, Sánchez-Díaz M, Antolin MC (2012) Effects of nitrogen source and water availability on stem carbohydrates and cellulosic bioethanol traits of alfalfa plants. *Plant Sci* 191–192:16–23.
- Fichot R, Laurans F, Monclus R, Moreau A, Pilate G, Brignolas F (2009) Xylem anatomy correlates with gas exchange, water-use efficiency and growth performance under contrasting water regimes: evidence from *Populus deltoides* × *Populus nigra* hybrids. *Tree Physiol* 29: 1537–1549.
- Fichot R, Barigah TS, Chamillard S, Le Thiec D, Laurans F, Cochard H, Brignolas F (2010) Common trade-offs between xylem resistance to cavitation and other physiological traits do not hold among unrelated *Populus deltoides* × *Populus nigra* hybrids. *Plant Cell Environ* 33: 1553–1568.
- Foster CE, Martin TM, Pauly M (2010) Comprehensive compositional analysis of plant cell walls (Lignocellulosic biomass) part I: lignin. *J Vis Exp* 37:e1745.
- Galbe M, Zacchi G (2002) A review of the production of ethanol from softwood. *Appl Microbiol Biotechnol* 59:618–628.
- Gaspar MJ, Alves A, Louzada JL, Morais J, Santos A, Fernandes C, Almeida MH, Rodrigues JC (2011) Genetic variation of chemical and mechanical traits of maritime pine (*Pinus pinaster* Aiton). Correlations with wood density components. *Ann For Sci* 68:255–265.
- Gonzalez-Martinez SC, Wheeler NC, Ersoz E, Nelson CD, Neale DB (2007) Association genetics in *Pinus taeda* L. I. Wood property traits. *Genetics* 175:399–409.
- Guerra FP, Wegrzyn JL, Sykes R, Davis MF, Stanton BJ, Neale DB (2013) Association genetics of chemical wood properties in black poplar (*Populus nigra*). *New Phytol* 197:162–176.
- Guet J, Fichot R, Lédée C, Laurans F, Cochard H, Delzon S, Bastien C, Brignolas F (2015) Stem xylem resistance to cavitation is related to xylem structure but not to growth and water-use efficiency at the within-population level in *Populus nigra* L. *J Exp Bot* 66: 4643–4652.

- Hamanishi ET, Raj S, Wilkins O, Thomas BR, Mansfield SD, Plant AL, Campbell MM (2010) Intraspecific variation in the *Populus balsamifera* drought transcriptome. *Plant Cell Environ* 33:1742–1755.
- Harholt J, Suttangkakul A, Scheller HV (2010) Biosynthesis of pectin. *Plant Physiol* 153:384–395.
- Harvey H, Van Den Driessche R (1997) Nutrition, xylem cavitation and drought resistance in hybrid poplar. *Tree Physiol* 17:647–654.
- Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* 8:R19.
- Hothorn T, Bretz F, Westfall P (2008) Simultaneous inference in general parametric models. *Biom J* 50:346–363.
- Hussey SG, Mizrahi E, Creux NM, Myburg AA (2013) Navigating the transcriptional roadmap regulating plant secondary cell wall deposition. *Front Plant Sci* 4:325.
- Janz D, Lautner S, Wildhagen H, Behnke K, Schnitzler JP, Rennenberg H, Fromm J, Polle A (2012) Salt stress induces the formation of a novel type of 'pressure wood' in two *Populus* species. *New Phytol* 194:129–141.
- Jin J, Tian F, Yang D-C, Meng Y-Q, Kong L, Luo J, Gao G (2016) PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 45(D1): D1040–D1045.
- Kilian J, Whitehead D, Horak J et al. (2007) The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J* 50: 347–363.
- Kim D, Perteza G, Trapnell C, Pimentel H, Kelley R, Salzberg S (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14:R36.
- Klash A, Ncube E, Toit B, du, Meincken M (2010) Determination of the cellulose and lignin content on wood fibre surfaces of eucalypts as a function of genotype and site. *Eur J For Res* 129:741–748.
- Kuang B, Zhao X, Zhou C et al. (2016) Role of UDP-glucuronic acid decarboxylase in xylan biosynthesis in *Arabidopsis*. *Mol Plant* 9: 1119–1131.
- Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559.
- Langfelder P, Horvath S (2012) Fast R functions for robust correlations and hierarchical clustering. *J Stat Softw* 46:111.
- Le Gall H, Philippe F, Domon J-M, Gillet F, Pelloux J, Rayon C (2015) Cell wall metabolism in response to abiotic stress. *Plants* 4:112–166.
- Lee C, O'Neill MA, Tsumuraya Y, Darvill AG, Ye Z-H (2007a) The irregular xylem9 mutant is deficient in xylan xylosyltransferase activity. *Plant Cell Physiol* 48:1624–1634.
- Lee C, Zhong R, Richardson EA, Himmelsbach DS, McPhail BT, Ye Z-H (2007b) The PARVUS gene is expressed in cells undergoing secondary wall thickening and is essential for glucuronoxylan biosynthesis. *Plant Cell Physiol* 48:1659–1672.
- Lee C, Teng Q, Huang W, Zhong R, Ye Z-H (2009) The Poplar GT8E and GT8F glycosyltransferases are functional orthologs of *Arabidopsis* PARVUS involved in glucuronoxylan biosynthesis. *Plant Cell Physiol*, 50:1982–1987.
- Lee C, Teng Q, Zhong R, Ye Z-H (2011) Molecular dissection of xylan biosynthesis during wood formation in Poplar. *Mol Plant* 4:730–747.
- Lenth R (2015) lsmeans: least-squares means. <http://CRAN.R-project.org/package=lsmeans> (31 August 2015, date last accessed).
- Li S, Bashline L, Zheng Y, Xin X, Huang S, Kong Z, Kim SH, Cosgrove DJ, Gu Y (2016) Cellulose synthase complexes act in a concerted fashion to synthesize highly aggregated cellulose in secondary cell walls of plants. *Proc Natl Acad Sci*, 113:11348–11353.
- Lionetti V, Francocci F, Ferrari S, Volpi C, Bellincampi D, Galletti R, D'Ovidio R, Lorenzo GD, Cervone F (2010) Engineering the cell wall by reducing de-methyl-esterified homogalacturonan improves saccharification of plant tissues for bioconversion. *Proc Natl Acad Sci USA* 107:616–621.
- Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550.
- Mansfield SD, Mooney C, Saddler JN (1999) Substrate and enzyme characteristics that limit cellulose hydrolysis. *Biotechnol Prog* 15: 804–816.
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17:10–12.
- Min D, Li Q, Jameel H, Chiang V, Chang H (2012) The cellulase-mediated saccharification on wood derived from transgenic low-lignin lines of black cottonwood (*Populus trichocarpa*). *Appl Biochem Biotechnol* 168:947–955.
- Moore JP, Vitré-Gibouin M, Farrant JM, Driouch A (2008) Adaptations of higher plant cell walls to water loss: drought vs desiccation. *Physiol Plant* 134:237–245.
- Muchero W, Guo J, DiFazio SP et al. (2015) High-resolution genetic mapping of allelic variants associated with cell wall chemistry in *Populus*. *BMC Genomics* 16:24.
- Nakano Y, Yamaguchi M, Endo H, Rejab NA, Ohtani M (2015) NAC-MYB-based transcriptional regulation of secondary cell wall biosynthesis in land plants. *Plant Physiol* 6:288.
- Nookaraju A, Pandey SK, Bae H-J, Joshi CP (2013) Designing cell walls for improved bioenergy production. *Mol Plant* 6:8–10.
- O'Brien TP, Feder N, McCully ME (1964) Polychromatic staining of plant cell walls by toluidine blue O. *Protoplasma* 59:368–373.
- Peña MJ, Zhong R, Zhou G-K, Richardson EA, O'Neill MA, Darvill AG, York WS, Ye Z-H (2007) *Arabidopsis* irregular xylem8 and irregular xylem9: implications for the complexity of glucuronoxylan biosynthesis. *Plant Cell* 19:549–563.
- Penning BW, Sykes RW, Babcock NC et al. (2014) Genetic determinants for enzymatic digestion of lignocellulosic biomass are independent of those for lignin abundance in a maize recombinant inbred population. *Plant Physiol* 165:1475–1487.
- Pinheiro J, Bates D, DebRoy S, Sarkar D, R Development Core Team (2015) nlme: linear and nonlinear mixed effects models. <http://CRAN.R-project.org/package=nlme> (31 August 2015, date last accessed).
- Plomion C, Leprovost G, Stokes A (2001) Wood formation in trees. *Plant Physiol* 127:1513–1523.
- Polle A, Janz D, Teichmann T, Lipka V (2013) Poplar genetic engineering: promoting desirable wood characteristics and pest resistance. *Appl Microbiol Biotechnol* 97:5669–5679.
- Porth I, Klapšte J, Skyba O et al. (2013) Genome-wide association mapping for wood characteristics in *Populus* identifies an array of candidate single nucleotide polymorphisms. *New Phytol* 200:710–726.
- Ranjan P, Yin T, Zhang X, Kalluri UC, Yang X, Jawdy S, Tuskan GA (2009) Bioinformatics-based identification of candidate genes from QTLs associated with cell wall traits in *Populus*. *BioEnergy Res* 3:172–182.
- R Development Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Rennie EA, Scheller HV (2014) Xylan biosynthesis. *Curr Opin Biotechnol* 26:100–107.
- Rogers LA, Dubos C, Surman C, Willment J, Cullis IF, Mansfield SD, Campbell MM (2005) Comparison of lignin deposition in three ectopic lignification mutants. *New Phytol* 168:123–140.
- Ruijter J, Ramakers C, Hoogaars W, Karlen Y, Bakker O, Van den Hoff M, Moorman A (2009) Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Res* 37:e45.
- Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27:863–864.
- Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9:671–675.

- Schreiber SG, Hacke UG, Hamann A, Thomas BR (2011) Genetic variation of hydraulic and wood anatomical traits in hybrid poplar and trembling aspen. *New Phytol* 190:150–160.
- Souza APD, Kamei CLA, Torres AF, Pattathil S, Hahn MG, Trindade LM, Buckeridge MS (2015) How cell wall complexity influences saccharification efficiency in *Miscanthus sinensis*. *J Exp Bot* 66:4351–4365.
- Srinivasasainagendra V, Page GP, Mehta T, Coulibaly I, Loraine AE (2008) CressExpress: a tool for large-scale mining of expression data from *Arabidopsis*. *Plant Physiol* 147:1004–1016.
- Studer MH, DeMartini JD, Davis MF, Sykes RW, Davison B, Keller M, Tuskan GA, Wyman CE (2011) Lignin content in natural *Populus* variants affects sugar release. *Proc Natl Acad Sci USA* 108:6300–6305.
- Tavares EQP, De Souza AP, Buckeridge MS (2015) How endogenous plant cell-wall degradation mechanisms can help achieve higher efficiency in saccharification of biomass. *J Exp Bot* 66:4133–4143.
- Teichmann T, Bolu-Arianto WH, Olbrich A, Langenfeld-Heyser R, Göbel C, Grzeganeck P, Feussner I, Hänsch R, Polle A (2008) GH3::GUS reflects cell-specific developmental patterns and stress-induced changes in wood anatomy in the poplar stem. *Tree Physiol* 28:1305–1315.
- Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M (2004) mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J* 37:914–939.
- Tuskan GA, DiFazio S, Jansson S et al. (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313:1596–1604.
- Van Acker R, Vanholme R, Storme V, Mortimer JC, Dupree P, Boerjan W (2013) Lignin biosynthesis perturbations affect secondary cell wall composition and saccharification yield in *Arabidopsis thaliana*. *Biotechnol Biofuels* 6:46.
- Viger M, Smith HK, Cohen D, Dewoody J, Trewin H, Steenackers M, Bastien C, Taylor G (2016) Adaptive mechanisms and genomic plasticity for drought tolerance identified in European black poplar (*Populus nigra* L.). *Tree Physiol* 36:909–928.
- Voelker SL, Lachenbruch B, Meinzer FC et al. (2010) Antisense down-regulation of 4CL expression alters lignification, tree growth, and saccharification potential of field-grown poplar. *Plant Physiol* 154:874–886.
- Wegrzyn JL, Eckert AJ, Choi M, Lee JM, Stanton BJ, Sykes R, Davis MF, Tsai C-J, Neale DB (2010) Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (*Populus trichocarpa*, Salicaceae) secondary xylem. *New Phytol* 188:515–532.
- Weih M, Polle A (2016) Editorial: ecological consequences of biodiversity and biotechnology in agriculture and forestry. *Front Plant Sci* 7:210.
- Wilkins O, Waldron L, Nahal H, Provart NJ, Campbell MM (2009) Genotype and time of day shape the *Populus* drought response. *Plant J* 60:703–715.
- van der Weijde T, Huxley LM, Hawkins S, Sembiring EH, Farrar K, Dolstra O, Visser RGF, Trindade LM (2016) Impact of drought stress on growth and quality of miscanthus for biofuel production. *GCB Bioenergy* 9:770–782.
- Xiao C, Anderson CT (2013) Roles of pectin in biomass yield and processing for biofuels. *Plant Biotechnol* 4:67.
- Xiao Z, Storms R, Tsang A (2004) Microplate-based filter paper assay to measure total cellulase activity. *Biotechnol Bioeng* 88:832–837.
- Yang JH, Wang H (2016) Molecular mechanisms for vascular development and secondary cell wall formation. *Front Plant Sci* 7:356.
- Ye Z-H, Zhong R (2015) Molecular control of wood formation in trees. *J Exp Bot* 66:4119–4131.
- Zhong R, Lee C, Ye Z-H (2010) Evolutionary conservation of the transcriptional network regulating secondary cell wall biosynthesis. *Trends Plant Sci* 15:625–632.
- Zhou G-K, Zhong R, Himmelsbach DS, McPhail BT, Ye Z-H (2007) Molecular characterization of PoGT8D and PoGT43B, two secondary wall-associated glycosyltransferases in Poplar. *Plant Cell Physiol*, 48:689–699.
- Zhou G, Taylor G, Polle A (2011) FTIR-ATR-based prediction and modeling of lignin and energy contents reveals independent intra-specific variation of these traits in bioenergy poplars. *Plant Methods* 7:1–10.