**TECHNICAL FOCUS**

**High-throughput microanalysis of large lignocellulosic sample sets by pyrolysis-gas chromatography/mass spectrometry**

Lorenz Gerber*, David Öhman*, Manoj Kumar*, Philippe Ranocha**, Deborah Goffner** and Björn Sundberg*

*aUmeå Plant Science Centre (UPSC), Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, Umeå, Sweden

**Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, Castanet-Tolosan, France

†Laboratoire de Recherche en Sciences Végétales, CNRS, Castanet-Tolosan, France

Correspondence
*Corresponding author, e-mail: lorenz.gerber@slu.se

Received 14 May 2015; revised 3 September 2015
doi:10.1111/ppl.12397

High-throughput analytical techniques to assess the chemistry of lignocellulosic plant material are crucial to plant cell-wall research. We have established an analytical platform for this purpose and demonstrated its usefulness with two applications. The system is based on analytical pyrolysis, coupled to gas chromatography/mass spectrometry – a technique particularly suited for analysis of lignocellulose. Automated multivariate-based data-processing methods are used to obtain results within a few hours after analysis, with an experimental batch of 500 analyzed samples. The usefulness of multivariate sample discrimination methods and hierarchical clustering of samples is demonstrated.

We have analyzed an *Arabidopsis* mutant collection consisting of 300 samples representing 31 genotypes. The mutant collection is presented through cluster analysis, based on chemotypic difference, with respect to wild type. Further, we have analyzed 500 thin sections from five biological replicate trees to create a spatial highly resolved profile of the proportions of syringyl-, guaiacyl- and p-hydroxyphenyl lignin across phloem, developing and mature wood in aspen. The combination of biologically easy to interpret information, the low demand of sample amount and the flexibility in sample types amenable to analysis makes this technique a valuable extension to the range of established high-throughput biomaterial analytical platforms.

**Introduction**

‘Phenomics’ is a general notion that applies to high-throughput methodologies for characterization of physical, chemical or biological phenotypes of plants (Kolukisaoglu and Thurow 2010). In research related to biomass production and lignocellulose properties, chemical characterization of cell walls is a central task. Often, large-scale approaches are crucial, which requires high-throughput methodologies. Moreover, the capability to handle small sample amounts is important for tissue-specific analysis, or in the studies of species

**Abbreviations** – G, guaiacyl; H, p-hydroxyphenyl; MCR-AR, multivariate curve resolution-alternating regression; OPLS-DA, orthogonal projection to latent structures-discriminant analysis; Py-GC/MS, pyrolysis-gas chromatography/mass spectrometry; S, syringyl.

†Present address: Faculty of Life Sciences, University of Manchester, Manchester, M13 9PT, UK

‡Present address: UMI 3189, ‘Environnement, Santé, Sociétés’, Faculté de Médecine secteur Nord, 51, Bd Pierre Dramard, 13344 Marseille, Cedex 15 France

© 2015 The Authors. Physiologia Plantarum published by John Wiley & Sons Ltd on behalf of Scandinavian Plant Physiology Society. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
that inherently yield small amounts of lignocellulose, such as Arabidopsis.

Traditional cell-wall analysis combines wet-chemical extraction, separation and fractionation steps prior to quantitative detection by gravimetry or an instrumental technique. One of the major concerns in wet-chemical approaches is to find reactions and conditions that allow a satisfactory specificity in the extraction of cell-wall polymers from the complex structure of the cell wall, despite the largely overlapping characteristics of their chemical bonds. This can sometimes be avoided by using highly specific enzymes instead of chemicals in the extraction process (Bauer et al. 2006). In general, wet-chemical approaches are tedious and difficult to automate for high-throughput approaches, but some breakthrough methods have been developed. A popular method for the determination of structural lignin polymer features, the ‘thioacidolysis’ protocol (Rolando et al. 1992), was recently adapted for increased throughput of microsamples (Robinson and Mansfield 2009). To identify specific differences in structural and compositional features in the non-cellulosic carbohydrate wall fraction, a combination of enzymatic pretreatment and MALDI-TOF analysis is a widespread approach capable of high sample throughput (Günl et al. 2011, Bauer 2012). Also based on enzymatic digestion is the PACE (polysaccharide analysis using carbohydrate gel electrophoresis) method, in which sugars are separated and quantified using gel electrophoresis (Goubet et al. 2002). A somewhat different approach to achieve high-throughput wet-chemical cell-wall analysis is described by Santoro et al. (2010), where the sample pretreatment, such as grinding, weighing and hydrolysis is automated by a dedicated lab robot.

As an alternative to isolation and analyses of individual wall components, cell-wall components can be analyzed without prior extraction and separation. One approach is to visualize defined epitopes using carbohydrate-binding modules and/or monoclonal antibody. High-throughput analysis of non-cellulosic wall carbohydrate polymers has been developed and demonstrated using ‘Comprehensive Micro-array Polysaccharide Profiling’ (Moller et al. 2007) as well as enzyme-linked immune absorbent assay (Pattathil et al. 2010). Another approach is to produce a chemical fingerprint (chemotype) of the whole cell-wall fraction, which is particularly advantageous in studies where an overall picture of cell-wall composition is wanted. There are several techniques that can provide cell-wall chemotypes. One possibility is to swell or dissolve the finely ground samples with suitable solvent system and measure them by 2D solution-state nuclear magnetic resonance (NMR) (Hedenström et al. 2009, Kim and Ralph 2010, Mansfield et al. 2012). The spectra can then be processed by multivariate data analysis tools that allow classification and characterization of the samples relative to each other (Hedenström et al. 2009). Although the level of structural detail on the polymer components is particularly high, the required sample amount for NMR is typically in the range of 20–100 mg, the sample preparation and analysis is not truly high-throughput, and quantitative analysis is problematic for many polymers due to peak overlap and the differences in relaxation rates between components. A frequently used spectroscopic method in plant cell-wall analysis is transmittance Fourier-transform infrared (FT-IR) spectroscopy. Here, sample preparation is less elaborate than for solution-state NMR, and rather quick acquisition times allow high throughput (Gorzsás and Sundberg 2014). FT-IR fingerprints are rich in information, but due to the overlap of characteristic bands from different polymers they do not translate so easily into chemical structure and composition, and quantitative estimates will be inaccurate (Gorzsás et al. 2011). However, chemical fingerprints obtained by FT-IR are well suited for sample classification using multivariate data analysis (Mouillé et al. 2003). Another technique that has been demonstrated to be useful for high-throughput analysis of complete cell-wall samples is Pyrolysis-Molecular-Beam Mass-Spectrometry. Here, pyrolysis of powder samples is used for reproducible sample fractionation, prior to mass spectrometric detection. The degradation products were shown to be specific for the different types of lignin (guaiacyl, syringyl and p-hydroxyphenyl) as well as pentose and hexose sugars (Tuskan et al. 1999). This approach provides semi-quantitative estimates of the above-mentioned cell-wall fractions, but the low sensitivity, due to the lack of chromatographic compound separation in combination with a scanning mass analyzer, is a limiting factor in, for example, screening and classification of mutant collections.

Pyrolysis-Gas Chromatography/Mass-Spectrometry (Py-GC/MS) is a well-established technique used in many different fields of polymer analysis (Moldoveanu 1998, Moldoveanu 2005). Specifically in wood sciences, this technique can be seen as a standard method for semi-quantitative determination of lignin structural features (Meier and Faix 1992, Alves et al. 2006). Also, GC/MS is often a standard instrumentation in plant cell-wall laboratories, and analytical pyrolysis can easily be retrofitted to enlarge the application envelope of an already existing instrument. Despite this, Py-GC/MS is not a widespread technique in molecular plant biology research. This may partly be due to the fact that data processing has been very tedious. A typical chromatogram from analytical biomass pyrolysis contains
normally 150–300 peaks. Until recently, the tools to process large set of chromatograms in a consistent and efficient way were limited. As mass-spectrometry-based high-throughput methods, such as metabolomics, become popular in life sciences, the number of available tools and adapted methods for data processing has increased. We recently demonstrated the usefulness of such a tool to process Py-GC/MS data in an efficient way (Gerber et al. 2012). We also developed a data analysis strategy to separate and classify chemical fingerprints that is useful in large-scale studies (Pinto et al. 2012). Here, we demonstrate high-throughput chemotyping of Arabidopsis cell-wall mutants by Py-GC/MS. Further, we show in a proof of concept study how Py-GC/MS and high-throughput data processing can be combined with spatially resolved microsampling to study aspects of wood development.

Materials and methods

Arabidopsis plant material

All mutants used in the screen are listed in Table S1, Supporting Information. irx5-4 (SALK_084627) was obtained from the Nottingham Arabidopsis Stock Centre (http://arabidopsis.info/), whereas irx6-2 (GABI_851H06) was obtained from GABI-Kat (Kleinboelting et al. 2012; http://www.gabi-kat.de/). Homozygous lines for all genotypes were identified by using gene-specific primers or by appropriate antibiotic selection. fra8 and irx9-2 were a kind gift from Dr Alan Marchant and c4h-3 and ccr1-6 from Dr Wout Boerjan. Seeds were sown on soil and plants grown for 4–5 weeks in 16 h, 24°C with 70% humidity. Dur-

ing growth, the position of the various genotypes and replicate plants was randomized. Basal 1–4 cm segments of 25–27 cm high stems were harvested, immediately frozen in liquid nitrogen and stored at −80°C until freeze drying. The freeze-dried stems were ground to powder in a ball-mill (MM400, Retsch, Germany; 2 min, 30 Hz, 80°C). The interface to the MS was kept for 3 min. The interface to vent away pyrolysate bleed of the sample remained on to vent away pyrolysate bleed of the sample remaining in the pyrolyzer oven. The pyrolysate was separated on a DB-5MS capillary column (30 m × 0.25 mm i.d., 0.25-μm-film thickness; J&W, Agilent Technologies AB, Sweden). The pyrolysis oven was set to 450°C, the interface to 340°C and the injector to 320°C. The injector was operated with a split ratio of 16:1 with helium as the carrier. After 1 min, the gas saver mode was switched on to vent away pyrolysate bleed of the sample remaining in the pyrolyzer oven. The pyrolysate was separated on a DB-5MS capillary column (30 m × 0.25 mm i.d., 0.25-μm-film thickness; J&W, Agilent Technologies AB, Sweden). The GC temperature program started at 40°C and increased with 32°C.min⁻¹ to 100°C, with 6°C.min⁻¹ to 118.75°C, with 15°C.min⁻¹ to 250°C and finally with 32°C.min⁻¹ to 320°C, where the temperature was kept for 3 min. The interface to the MS was kept at 280°C. The mass spectrometer scanned the range from m/z 35 to m/z 250, resulting in a scan rate of 6.22 scans s⁻¹.

Pyrolysis-gas chromatography/mass spectrometry and raw data processing

The analytical setup consisted of an oven pyrolyser equipped with an autosampler (PY-2020iD and AS-1020E, Frontier Labs, Japan) connected to a GC/MS system (Agilent, 7890A/5975C, Agilent Technologies AB, Sweden). The pyrolysis oven was set to 450°C, the interface to 340°C and the injector to 320°C. The injector was operated with a split ratio of 16:1 with helium as the carrier. After 1 min, the gas saver mode was switched on to vent away pyrolysate bleed of the sample remaining in the pyrolyzer oven. The pyrolysate was separated on a DB-5MS capillary column (30 m × 0.25 mm i.d., 0.25-μm-film thickness; J&W, Agilent Technologies AB, Sweden). The GC temperature program started at 40°C and increased with 32°C.min⁻¹ to 100°C, with 6°C.min⁻¹ to 118.75°C, with 15°C.min⁻¹ to 250°C and finally with 32°C.min⁻¹ to 320°C, where the temperature was kept for 3 min. The interface to the MS was kept at 280°C. The mass spectrometer scanned the range from m/z 35 to m/z 250, resulting in a scan rate of 6.22 scans s⁻¹.

Raw data processing

Raw data files were converted to NetCDF format in Agilent ChemStation Data Analysis (Version E.02.00.493). All further data processing was done as described elsewhere (Gerber et al. 2012). Briefly, data preprocessing involved smoothing and alignment of all chromatograms. For performance reasons, and as a preparative step for background correction, chromatograms were then divided into windows of 200 to 450 scans. Prior to applying Multivariate Curve-Resolution by Alternate Regression (MCR-AR), for resolving (deconvoluting) the chromatographic and mass spectral profiles, every window was baseline corrected by linear interpolating and subsequent subtracting of the area under the line between the first and the last data point in a processing window.
The MCR-AR processing step yielded a table with integrals per peak and sample and a peak table with mass spectra.

**Data analysis Arabidopsis samples**

Data analysis was done according to Pinto et al. (2012) using the programming environment R (R Core Team 2013). Briefly, for each mutant line, 1000 orthogonal projection to latent structures-discriminant analysis (OPLS-DA) models were built. For each model a number of wild types equal to the available mutant specimens were sampled randomly from the Columbia or Wassilewskija wild-type population. To achieve comparable models, one orthogonal component was calculated for each model. In each round, the loading vector was summed and in the end averaged to the number of calculated models. The number of misclassified mutant samples was determined in each round and summed. The resulting averaged loading vectors were then subjected to hierarchical clustering by Ward’s method (Ward 1963) and visualized in a heatmap (R function `heatmap()`).

**Data analysis Populus samples**

The processed data were classified into S-, G- and H-type lignin according to the base peak method described in Gerber et al. (2012). The values in each data series were normalized to the total integrated signal in each measured sample. Data series of each tree were aligned according to visual references obtained during sectioning and plotted using R (R Core Team 2013).

**Results and discussion**

**Multivariate curve resolution enables high-throughput data processing**

To aim for a high-throughput application, the heating rate in the temperature program of the GC was set to optimize for short, but still reproducible, retention times. We used run times of 19 min, giving a capacity of 300 samples in 7 days under continuous operation. To account for unavoidable drift and sensitivity fluctuations, the run order of the samples was fully randomized. The chromatogram resulting from each run then has to be data processed, i.e., smoothed, aligned, peak-picked and peak-integrated.

With traditional data-processing tools, these steps imposed a major problem for high-throughput Py-GC/MS applications because it involved individual processing of each chromatogram. Moreover, peaks are usually not well separated with the short run time required for high-throughput applications, and a deconvolution step is required to allow for reliable peak integration. Matching corresponding peaks from different runs then becomes complicated due to the unavoidable ambiguity of a deconvoluted chromatogram and normally requires human interaction. To overcome these bottlenecks, MCR-AR was adapted to Py-GC/MS data handling (Gerber et al. 2012). In the MCR-AR approach, chromatograms are processed in parallel. Initially, about 40 chromatograms with the largest variation in the dataset are selected by principal component analysis and processed by MCR-AR. This yields deconvoluted mass spectra and peak retention times that are used to automatically fit the peaks of the remaining chromatograms according to their mass spectra and retention times. As the initial MCR-AR process takes typically 2 h, the calculation of remaining chromatograms is done in less than a minute. Therefore, processing time is independent of the number of samples to be processed. There is no upper limit of samples that can be processed together, but the long-time drift of the chromatographic system will induce too much variation in peak retention for accurate peak identification if too many samples are included in one experiment. In our experience, experiments with more than 500 samples should preferentially be processed in split sets.

After processing of the chromatograms, peaks are automatically classified and integrated. Classification defines peaks into syringyl (S), guaiacyl (G), and \(p\)-hydroxyphenyl (H) lignin, carbohydrates (C) and unknowns (U), according to the highest abundant m/z channel, as described in Gerber et al. (2012). Summation of peaks from the same class allows for estimation of, for example, S to G ratios, total lignin or total carbohydrate fraction. It should be noted, however, that automated classification of the pyrolytic degradation products is not always transferable between different sample types and species. For example, in grasses pyrolytic degradation products of \(p\)-coumarate and ferulates result in overestimation of H- and G-type lignin (del Rio et al. 2012). Further pyrolysis of polysaccharides can yield phenol (Moldoveanu 1998), a degradation product that in literature is often used as a main contributor for calculation of H-type lignin (Faix et al. 1990). Thus, automated classification should ideally be adjusted for the species under investigation.

**Data analysis based on OPLS-DA is used to identify and classify mutant genotypes with modified chemotypes**

To show the usefulness of Py-GC/MS in cell-wall research, we performed a screen of Arabidopsis T-DNA mutants, representing 31 genotypes of Colombia and
Wassilewskija ecotypes, denoted DG1 to 35, described in Table S1. Selected mutants were disrupted in genes homologous to genes that showed increased expression during tracheary element differentiation in *Zinnia* cell cultures, and therefore putatively involved in secondary cell-wall biosynthesis (Pesquet et al. 2005). At the time of selection, all mutants were uncharacterized, but during the course of this study, DG7 and DG8 were shown to be deficient in xylem secondary cell walls as a result of a mutation in *wat1*, a gene encoding an auxin transporter (Ranocha et al. 2010, Ranocha et al. 2013). A number of positive controls were included, representing earlier characterized cell-wall mutants. These were *ccr1-6*, coding for cinnamoyl-CoA reductase (Mir Derikvand et al. 2008) and *c4h-3*, coding for cinnamyl-4-hydroxylase (Ruegger and Chapple 2001), both involved in lignin biosynthesis; *irx9-2*, coding for a family 43 glycosyltransferase (Wu et al. 2010) and *ira8*, coding for family 47 glycosyltransferase (Zhong et al. 2005), both involved in xylan biosynthesis; and *irx5-4*, coding for *CESA4* (Brown et al. 2005) and *irx6-2*, coding for a COBRA-LIKE4 (COBL4) enzyme (Li et al. 2003, Sindhu et al. 2007), both affecting cellulose biosynthesis. Moreover, 66 wild-type controls were included to have a good coverage of the variation in the wild-type population. Altogether the whole experiment included 313 samples collected from the basal part of the inflorescence stem. Such samples typically yielded chromatograms with 227 resolved peaks after Py-GC/MS analysis. To take full advantage of the information from all peaks in the chromatogram, we analyzed the data set with the multivariate discrimination method ‘orthogonal projection to latent structures by partial least squares – discriminant analysis’ (OPLS-DA) (Bylesjö et al. 2006). Like other multivariate regression methods, such as PLS, OPLS seeks to describe a relation between the data matrix (in this case, peak areas of all defined peaks) and the class descriptors (in this case, classification of a sample as wild type or mutant). The advantage of OPLS-DA as compared with other multivariate regression methods is that it decomposes the described variation into only one discriminant component used for predicting the difference between the wild-type and mutant sample, and one or several orthogonal components to describe variation within the replicates of wild-type and mutant samples (Bylesjö et al. 2006). The predictive component resulting from an OPLS-DA analysis can be expressed as loadings vectors that describe the chromatographic peaks of importance for defining the sample classification.

A common, and sometimes unavoidable, problem for comparisons in large-scale studies is systematic variation that makes it complicated to combine all data into one global dataset. Systematic variation can be due to drift of the instrument during analysis, but it can also arise when material from different growth batches are used, e.g., due to unavoidable differences in environmental conditions for different growth batches, or due to the use of different ecotypes (as in this example). To circumvent this problem, we developed a strategy based on hierarchical clustering of the loading vectors from bootstrapped and class-balanced OPLS-DA models between one mutant genotype and wild-type samples. The theoretical background to the data analysis is described in detail elsewhere (Pinto et al. 2012). Briefly, a large number of OPLS-DA models between the replicate samples of one mutant genotype and an equal number of randomly chosen wild types of the corresponding ecotype are iteratively calculated (bootstrapped). In this case, we ran 1000 iterations, each with a different set of replicate wild-type controls. The predictive loading vectors describing the OPLS-DA model from all the 1000 calculations were summed and finally averaged. Bootstrapping OPLS-DA models is in this case beneficial because it yields class-balanced models (i.e., each model is built on equal number of biological replicates plants in each class) and number of misclassified predictions.

Loading vectors were obtained for OPLS-DA models of the 31 genotypes included in our example by the above-described approach. Together these loading vectors constitute a data matrix, which is then reordered by hierarchical cluster analysis (HCA) based on Ward’s method (Ward 1963) and visualized in a heatmap (Fig. 1). Genotypes with a similar loading fingerprint (i.e., similar difference to wild type) indicate a similar chemotype and should be clustered together. To evaluate whether the reordering of the heatmap is meaningful, we added the peak classification labels, according to Gerber et al. (2012) (i.e., Carbohydrate, Syringyl, Guaiacyl, p-Hydroxyphenyl, other Phenolic, Known spectra, Unknown spectra) into the figure. It was found that peaks with the same classification in general appear together in a group (Fig. 1). In a similar way, known mutants can be used to validate the HCA reordering of the rows. *irx6-2* and *irx5-4*, which are both cellulose-related mutants (*CESA4*, *COBL4*) and *irx9-2* and *ira8*, two glucurono-xylan backbone-related genes, are clustered together. The pattern of the loadings fingerprint for these four mutants shows a general decrease in many carbohydrate-related peaks (blue color), whereas the lignin-related peaks increased in concert (red color). The increase in lignin is a normalization effect. The phenylpropanoid pathway mutants, *ccr1-6* and *c4h-3*, are also clustered as direct neighbors with loading fingerprints (color patterns) opposite to the above-described; i.e., a
decrease in lignin-related peaks, and as a result of normalization, an increase in carbohydrate-related peaks. It can also be noted that the two wat1 alleles (DG07 and DG08) clustered together, despite being in different genetic backgrounds. It should, however, be pointed out that the clustering in the heatmap is dependent on both the loading fingerprint and the strength of each loading. Therefore, mutants with a similar chemotype but of different strength will not necessarily cluster together. To identify genotypes most different from wild type,
we used the number of misclassifications of the 1000 iterative OPLS models performed for each comparison between genotype and wild type as a metric for the predictive strength of the bootstrapped OPLS-DA models. Model misclassifications are defined as the number of mutant samples, having discriminant component scores that position them on the wild-type side in the score plot. The lower the number of misclassifications, the more likely the mutant differs from wild type.

All the earlier described mutants showed strong chemotypes with zero misclassifications, except for irx5-4 with three misclassifications. Eight of the DG mutants showed less than 10 misclassifications, and were considered having a strong chemotype, as compared to wild type. Two of these were the wt1 mutants, described earlier. The other six mutants (SKS17, AGO4, nodulin MtN21, peroxidase superfamily protein, ASD2 and CB5-D) are, however, not previously described and can be considered as strong candidates to have a mutation in important genes in the regulation or biosynthesis of cell walls. None of these mutants were significantly aberrant in overall growth or had any obvious anatomical phenotype, and therefore would not have been detected by any visual screening approach. It should also be noted that the T-DNA mutants with an almost similar chemotype to wild type only resulted in between 66 and 90 misclassifications from the 1000 comparisons (i.e., DG1, DG3, DG4, DG11 and DG12). This indicates a high risk of false positives using a more traditional approach, comparing only five wild-type and mutant sample each, and shows the advantage of high-throughput analysis allowing for a large number wild-type controls to cover the full variation in wild type.

In this case study, we decided to analyze the samples without prior removal of soluble extractives. Therefore, the chemotypes will not only reflect differences in the insoluble cell wall but could also reflect differences in secondary metabolites, sugars, oligolignols or other soluble compounds in the cell. The same approach as described above can be applied on samples where extractives have been removed, if the study is more focused on insoluble wall components. But extracting a large number of samples in a reproducible manner is a very tedious process and such approach will be less high throughput.

**Pattern of syringyl, guaiacyl and \( p \)-Hydroxyphenyl lignin during wood development**

With Py-GC/MS, small sample amounts can be analyzed. We have obtained reliable data from samples down to 1 \( \mu \)g. Thus, microanalysis can be performed on, for example, dissected tissues. In the case of our analytical setup, different types of samples (powder, solid or liquid) can be applied to the pyrolyzer in deactivated stainless steel sample cups that are handled by an autosampler. Chemotypes can be evaluated using the multivariate approach described above, or quantitative estimates of cell-wall polymers can be obtained by peak integration. The latter is particularly interesting when the aim is to study lignin profiles, as Py-GC/MS separates the different lignin types. Here, we provide an example of lignin profiling in micro-dissected tissues obtained across wood forming tissues.

The analysis was done on cryo-dissected longitudinal tangential sections obtained from the phloem across the cambium and to the annual ring in samples collected from field-grown aspen trees. This sampling approach has previously been applied to visualize, for example, auxin distribution or gene expression across developing wood with micro-analytical techniques (Uggla et al. 1996, Hertzberg et al. 2001) and is described in detail elsewhere (Uggla and Sundberg 2001). Each tangential section was 10 \( \mu \)m thick and weighed from a few micrograms (cambium) up to several hundred micrograms (mature xylem). The sections were freeze-dried in custom-made paper-lined sample boxes to prevent them from sticking to the container. With the high-throughput data processing, it was feasible to analyze 100 samples per section series for five replicate trees in one experiment (total of 500 samples).

After removal of obvious outliers and curve smoothing, the lignin composition is visualized across wood forming tissues with a high agreement among the replicate trees (Fig. 2). Whereas the proportion of the S-type lignin type follows an even increase up to ca 11%, the G-lignin type reaches a plateau at about 700 \( \mu \)g. Thus, microanalysis can be performed on, for example, dissected tissues. In the case of our analytical setup, different types of samples (powder, solid or liquid) can be applied to the pyrolyzer in deactivated stainless steel sample cups that are handled by an autosampler. Chemotypes can be evaluated using the multivariate approach described above, or quantitative estimates of cell-wall polymers can be obtained by peak integration. The latter is particularly interesting when the aim is to study lignin profiles, as Py-GC/MS separates the different lignin types. Here, we provide an example of lignin profiling in micro-dissected tissues obtained across wood forming tissues.

The analysis was done on cryo-dissected longitudinal tangential sections obtained from the phloem across the cambium and to the annual ring in samples collected from field-grown aspen trees. This sampling approach has previously been applied to visualize, for example, auxin distribution or gene expression across developing wood with micro-analytical techniques (Uggla et al. 1996, Hertzberg et al. 2001) and is described in detail elsewhere (Uggla and Sundberg 2001). Each tangential section was 10 \( \mu \)m thick and weighed from a few micrograms (cambium) up to several hundred micrograms (mature xylem). The sections were freeze-dried in custom-made paper-lined sample boxes to prevent them from sticking to the container. With the high-throughput data processing, it was feasible to analyze 100 samples per section series for five replicate trees in one experiment (total of 500 samples).

After removal of obvious outliers and curve smoothing, the lignin composition is visualized across wood forming tissues with a high agreement among the replicate trees (Fig. 2). Whereas the proportion of the S-type lignin type follows an even increase up to ca 11%, the G-lignin type reaches a plateau at about 700 \( \mu \)g. Thus, microanalysis can be performed on, for example, dissected tissues. In the case of our analytical setup, different types of samples (powder, solid or liquid) can be applied to the pyrolyzer in deactivated stainless steel sample cups that are handled by an autosampler. Chemotypes can be evaluated using the multivariate approach described above, or quantitative estimates of cell-wall polymers can be obtained by peak integration. The latter is particularly interesting when the aim is to study lignin profiles, as Py-GC/MS separates the different lignin types. Here, we provide an example of lignin profiling in micro-dissected tissues obtained across wood forming tissues.

The analysis was done on cryo-dissected longitudinal tangential sections obtained from the phloem across the cambium and to the annual ring in samples collected from field-grown aspen trees. This sampling approach has previously been applied to visualize, for example, auxin distribution or gene expression across developing wood with micro-analytical techniques (Uggla et al. 1996, Hertzberg et al. 2001) and is described in detail elsewhere (Uggla and Sundberg 2001). Each tangential section was 10 \( \mu \)m thick and weighed from a few micrograms (cambium) up to several hundred micrograms (mature xylem). The sections were freeze-dried in custom-made paper-lined sample boxes to prevent them from sticking to the container. With the high-throughput data processing, it was feasible to analyze 100 samples per section series for five replicate trees in one experiment (total of 500 samples).

After removal of obvious outliers and curve smoothing, the lignin composition is visualized across wood forming tissues with a high agreement among the replicate trees (Fig. 2). Whereas the proportion of the S-type lignin type follows an even increase up to ca 11%, the G-lignin type reaches a plateau at about 700 \( \mu \)g. Thus, microanalysis can be performed on, for example, dissected tissues. In the case of our analytical setup, different types of samples (powder, solid or liquid) can be applied to the pyrolyzer in deactivated stainless steel sample cups that are handled by an autosampler. Chemotypes can be evaluated using the multivariate approach described above, or quantitative estimates of cell-wall polymers can be obtained by peak integration. The latter is particularly interesting when the aim is to study lignin profiles, as Py-GC/MS separates the different lignin types. Here, we provide an example of lignin profiling in micro-dissected tissues obtained across wood forming tissues.

The analysis was done on cryo-dissected longitudinal tangential sections obtained from the phloem across the cambium and to the annual ring in samples collected from field-grown aspen trees. This sampling approach has previously been applied to visualize, for example, auxin distribution or gene expression across developing wood with micro-analytical techniques (Uggla et al. 1996, Hertzberg et al. 2001) and is described in detail elsewhere (Uggla and Sundberg 2001). Each tangential section was 10 \( \mu \)m thick and weighed from a few micrograms (cambium) up to several hundred micrograms (mature xylem). The sections were freeze-dried in custom-made paper-lined sample boxes to prevent them from sticking to the container. With the high-throughput data processing, it was feasible to analyze 100 samples per section series for five replicate trees in one experiment (total of 500 samples).

After removal of obvious outliers and curve smoothing, the lignin composition is visualized across wood forming tissues with a high agreement among the replicate trees (Fig. 2). Whereas the proportion of the S-type lignin type follows an even increase up to ca 11%, the G-lignin type reaches a plateau at about 700 \( \mu \)g. Thus, microanalysis can be performed on, for example, dissected tissues. In the case of our analytical setup, different types of samples (powder, solid or liquid) can be applied to the pyrolyzer in deactivated stainless steel sample cups that are handled by an autosampler. Chemotypes can be evaluated using the multivariate approach described above, or quantitative estimates of cell-wall polymers can be obtained by peak integration. The latter is particularly interesting when the aim is to study lignin profiles, as Py-GC/MS separates the different lignin types. Here, we provide an example of lignin profiling in micro-dissected tissues obtained across wood forming tissues.
Fig. 2. Relative Syringyl-, Guayacyl- and \( p \)-Hydroxyphenyl- type lignin content from phloem, across the developing wood to the annual ring in five replicate aspen trees (A–E). Each symbol represents an independent sample of a 10 µm tangential sections analyzed by pyrolysis-gas chromatography/mass spectrometry. The values are expressed in percentage of the total integrated peak area per sample.
secondary cell wall (Fukushima and Terashima 1990). However, it has also been found that both S- and G-types of monolignols are present as extractable compounds, usually in the form of glucosides, in the cambial zone (Fukushima and Terashima 1990). Our samples were not extracted to remove soluble compounds before analysis. Hence, the presented data should not be used to experimentally confirm the temporal and spatial specific deposition of S-, G- and H-lignin types into insoluble lignin at the very early stage of lignification. However, the data suggests that there is no distinct difference in the temporal regulation of the presence of S- and G-types of monolignols during the wood forming process (Fig. 3). During the polymerization of insoluble lignin macromolecules, however, the monolignol precursors may be utilized in a controlled fashion to give rise to the temporal regulation of S- and G-enrichment in the lignin, as described above. Another possibility is that the spatial distribution of lignin monomers differs, for example, between the symplast and apoplast.

Conclusions

Taken together, Py-GC/MS is a well-established analytical technique for a wide range of lignocellulose samples. The application of multivariate curve resolution in Py-GC/MS data analysis now allows for truly high-throughput experiments. Py-GC/MS can be used for quantitative analysis of cell-wall components of microsamples and is particularly suited for lignin components as was demonstrated here by visualizing their presence during wood development. Another application in Py-GC/MS data analysis is the use of multivariate tools to classify chemotypes and to mitigate the normally unavoidable systematic variation in large sample sets. In our example, we could identify novel cell-wall mutants in an Arabidopsis T-DNA mutants collection. However, the approach could also be used in a wide spectrum of investigations, e.g. to identify the impact of biotic and abiotic parameters on cell-wall chemotype in physiological studies and to screen for biological diversity. With these breakthrough applications in Py-GC/MS data analysis, the technology is now unlocked for a much more versatile use in plant biology research.

Authors’ contributions

Design of experiments was done by L. G., B. S., D. G. Experiments were performed by L. G., M. K., D. Ö., and
Data analysis was done by L. G. The manuscript was written by L. G. and B. S.

Acknowledgements – This work was supported by grants from FORMAS (Func-Fiber/Bioimprove – centre of excellence in wood science), Vetenskapsrådet, the Swedish Energy Agency, EU Programme Renewall (FP7/2007–2013), VINNOVA, the French National Research Agency (ANR Génoplante WALLTALK ANR-07-GPLA-014) and Bio4Energy, the Swedish Programme for renewable energy. The authors would like to thank Amandine Freydier for her invaluable help in identifying the homozygous Arabidopsis lines.

References

Alves A, Schwanninger M, Pereira H, Rodrigues J (2006) Analytical pyrolysis as a direct method to determine the lignin content in wood. Part 1: comparison of pyrolysis lignin with Klason lignin. J Anal Appl Pyrolysis 76: 209–213

Bauer S (2012) Mass spectrometry for characterizing plant cell wall polysaccharides. Front Plant Sci 3: 45

Bauer S, Vasu P, Persson S, Mort AJ, Somerville CR (2006) Development and application of a suite of polysaccharide-degrading enzymes for analyzing plant cell walls. Proc Natl Acad Sci USA 103: 11417–11422

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

Bylesjö M, Rantalainen M, Cloarec O, Nicholson JK, Holmes E, Trygg J (2006) OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. J Chemom 20: 341–351

Coleman HD, Park J-Y, Nair R, Chapple C, Mansfield SD (2008) RNAi-mediated suppression of 3‘-coumaroyl-CoA 3′-hydroxylase in hybrid poplar impacts lignin deposition and soluble secondary metabolism. Proc Natl Acad Sci USA 105: 4501–4506

de Rio J, Rencoret J, Prinsen P, Martinez A, Ralph J, Gutiérrez A (2012) Structural Characterization of Wheat Straw Lignin as Revealed by Analytical Pyrolysis, 2D-NMR, and Reductive Cleavage Methods. J Agric Food Chem 60: 5922–5935

Donaldson LA (1994) Mechanical constraints on lignin deposition during lignification. Wood Sci Technol 28: 111–118

Donaldson LA (2001) Lignification and lignin topochemistry – an ultrastructural view. Phytochemistry 57: 859–873

Faix O, Maier D, Fortman I (1990) Thermal degradation products of wood. Holz als Roh- und Werkstoff 48: 281–285

Fukushima K, Terashima N (1990) Heterogeneity in formation of lignin. XIII. Formation of p-hydroxyphenyl lignin in various hardwoods visualized by microautoradiography. J Wood Chem Technol 10: 413–433

Gerber L, Eliasson M, Trygg J, Moritz T, Sundberg B (2012) Multivariate curve resolution provides a high-throughput data processing pipeline for pyrolysis-gas chromatography/mass spectrometry. J Anal Appl Pyrolysis 95: 95–100

Gorzás A, Sundberg B (2014) Chemical fingerprinting of Arabidopsis using Fourier transform infrared (FT-IR) spectroscopic approaches. In: Sanchez-Serrano JJ, Salinas J (eds) Arabidopsis Protocols. Humana Press, New York, pp 317–352

Gorzás A, Stenlund H, Persson P, Trygg J, Sundberg B (2011) Cell-specific chemotyping and multivariate imaging by combined FT-IR microspectroscopy and orthogonal projections to latent structures (OPLS) analysis reveals the chemical landscape of secondary xylem. Plant J 66: 903–914

Goubet F, Jackson P, Deere MJ, Dupree P (2002) Polysaccharide analysis using carbohydrate gel electrophoresis: a method to study plant cell wall polysaccharides and polysaccharide hydrolases. Anal Biochem 300: 53–68

Günl M, Kraemer F, Pauly M (2011) Oligosaccharide mass profiling (OLIMP) of cell wall polysaccharides by MALDI-TOF/MS. In: Popper ZA (ed) The Plant Cell Wall. Humana Press, pp 43–54

Hedenström M, Wiklund-Lindström S, Oman T, Lu F, Gerber L, Schatz P, Sundberg B, Ralph J (2009) Identification of lignin and polysaccharide modifications in Populus wood by chemometric analysis of 2D NMR spectra from dissolved cell walls. Mol Plant 2: 933–942

Hertzberg M, Aspeborg H, Schrader J, Andersson A, Erlandsson R, Blomqvist K, Bhalerao R, Uhlen M, Teeri TT, Lundeberg J, Sundberg P, Nilsson P, Sundberg G (2001) A transcriptional roadmap to wood formation. Proc Natl Acad Sci USA 98: 14732–14737

Kim H, Ralph J (2010) Solution-state 2D NMR of ball-milled plant cell wall gels in DMSO-d6/pyridine-d5. Org Biomol Chem 8: 576–591

Kleinboelting N, Huep G, Kloetgen A, Viehöver P, Weisshaar B (2012) GABI-Kat SimpleSearch: new features of the Arabidopsis thaliana T-DNA mutant database. Nucleic Acids Res 40: D1211–D1215

Kolukisaoglu Ü, Thurow K (2010) Future and frontiers of automated screening in plant sciences. Plant Sci 178: 476–484

Li Y, Qian Q, Zhou Y, Yan M, Sun L, Zhang M, Fu Z, Wang Y, Han B, Pang X, et al. (2003) BRITTLE CULM1, which encodes a COBRA-like protein, affects the mechanical properties of rice plants. Plant Cell 15: 2020–2031
Mansfield SD, Kim H, Lu F, Ralph J (2012) Whole plant cell wall characterization using solution-state 2D NMR. Nat Protocols 7: 1579–1589

Meier D, Faix O (1992) Pyrolysis-gas chromatography–mass spectrometry. In: Lin SY, Dence CW (eds) Methods in Lignin Chemistry. Springer, Berlin Heidelberg, pp 177–199

Mir Derikvand M, Sierra J, Ruel K, Pollet B, Do C-T, Thévenin J, Buffard D, Jouanin L, Lapierre C (2008) Redirection of the phenylpropanoid pathway to feruloyl malate in Arabidopsis mutants deficient for cinnamoyl-CoA reductase 1. Planta 227: 943–956

Moldoveanu SC (1998) Introduction and nomenclature; Analytical pyrolysis of polymeric carbohydrates. In: Moldoveanu SC (ed) Analytical Pyrolysis of Natural Organic Polymers, Vol. 3–7. Elsevier, Amsterdam, pp 291–297

Moldoveanu SC (2005) Polymers with unsaturated carbon chain backbone. In: Moldoveanu SC (ed) Analytical Pyrolysis of Synthetic Organic Polymers. Elsevier, Amsterdam, pp 439–461

Moller I, Sørensen I, Bernal AJ, Blaukopf C, Lee K, Obro J, Pettolino F, Roberts A, Mikkelsen JD, Knox JP, Bacic A, Willats WG (2007) High-throughput mapping of cell-wall polymers within and between plants using novel microarrays. Plant J 50: 1118–1128

Mouille G, Robin S, Lecomte M, Pagant S, Höfte H (2003) Moller I, Sørensen I, Bernal AJ, Blaukopf C, Lee K, Obro J, Pettolino F, Roberts A, Mikkelsen JD, Knox JP, Bacic A, Willats WG (2007) High-throughput mapping of cell-wall polymers within and between plants using novel microarrays. Plant J 50: 1118–1128

Mouille G, Robin S, Lecomte M, Pagant S, Höfte H (2003) Classification and identification of Arabidopsis cell wall mutants using Fourier-Transform InfraRed (FT-IR) microspectroscopy. Plant J 35: 393–404

Pattathil S, O'Neill M, York W, Hahn MG (2010) A comprehensive toolkit of plant cell wall glycan-directed monoclonal antibodies. Plant Physiol 153: 514–525

Pinto RC, Gerber L, Eliasson M, Sundberg B, Trygg J (2012) Strategy for minimizing between-study variation of large-scale phenotypic experiments using multivariate analysis. Anal Chem 84: 8675–8681

R Core Team (2013) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at http://www.R-project.org/

Ranocha P, Denancé N, Vanholme R, Freydier A, Martinez Y, Hoffmann L, Köhler L, Pouzet C, Renou J-P, Sundberg B, Boerjan W, Goffner D (2010) Walls are thin in Medicago truncatula NODULIN21, is a tonoplast-localized protein required for secondary wall formation in fibers. Plant J 63: 469–483

Ranocha P, Dima O, Nagy R, Felten J, Corratgé-Faillie C, Novák O, Morreel K, Lacombe B, Martinez Y, Pfrunder S, Jin X, Renou JP, Thibaud JB, Ljung K, Fischer U, Martinova E, Boerjan W, Goffner D (2013) Arabidopsis WAT1 is a vacuolar auxin transport facilitator required for auxin homeostasis. Nat Commun 4: 2625

Robinson AR, Mansfield SD (2009) Rapid analysis of poplar lignin monomer composition by a streamlined thioacidolysis procedure and near-infrared reflectance-based prediction modeling. Plant J 58: 706–714

Rolando C, Monties B, Lapiere C (1992) Thioacidolysis. In: Lin SY, Dence CW (eds) Methods in Lignin Chemistry. Springer, Berlin Heidelberg, pp 334–349

Ruegger M, Chapple C (2001) Mutations that reduce sinapoylmalate accumulation in Arabidopsis thaliana define loci with diverse roles in phenylpropanoid metabolism. Genetics 159: 1741–1749

Santoro N, Cantu S, Tornqvist C-E, Falbel T, Bolivar J, Patterson S, Pauly M, Walton J (2010) A high-throughput platform for screening milligram quantities of plant biomass for lignocellulose digestibility. Bioenergy Res 3: 93–102

Sindhu A, Langewisch T, Olek A, Multani DS, McCann MC, Vermerris W, Carpita NC, Johal G (2007) Maize Brittle stalk2 encodes a COBRA-like protein expressed in early organ development but required for tissue flexibility at maturity. Plant Physiol 145: 1444–1459

Terashima N (1990) A new mechanism for formation of a structurally ordered protolignin macromolecule in the cell wall of tree xylem. J Pulp Paper Sci 16: J150–J155

Terashima N, Awano T, Takabe K, Yoshida M (2004) Formation of macromolecular lignin in ginkgo xylem cell walls as observed by field emission scanning electron microscopy. C R Biol 327: 903–910

Tuskan G, West D, Bradshaw H, Neale D, Sewell M, Wheeler N, Megraw B, Jech K, Wiselogel A, Evans R, Elam C, Davis M, Dinus R (1999) Two high-throughput techniques for determining wood properties as part of a molecular genetics analysis of hybrid poplar and loblolly pine. Appl Biochem Biotechnol 77: 55–65

Ugglia C, Sandberg B (2001) Sampling of cambial region tissues for high resolution analysis. In: Chaffey NJ (ed) Wood Formation in Trees: Cell and Molecular Biology Techniques. CRC Press, Taylor & Francis Group, LLC, London and New York, pp 215–228

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

Ugglia C, Mellerowicz EJ, Sandberg B (1998) Indole-3-acetic acid controls cambial growth in Scots pine by positional signaling. Plant Physiol 117: 113–121
Ward JH Jr (1963) Hierarchical grouping to optimize an objective function. J Am Stat Assoc 58: 236–244
Wu A-M, Hörnblad E, Voxeur A, Gerber L, Rihouey C, Lerouge P, Marchant A (2010) Analysis of the Arabidopsis IRX9/IRX9-L and IRX14/IRX14-L pairs of glycosyltransferase genes reveals critical contributions to biosynthesis of the hemicellulose glucuronoxylan. Plant Physiol 153: 542–554
Zhong R, Peña MJ, Zhou G-K, Nairn CJ, Wood-Jones A, Richardson EA, Morrison WH, Darvill AG, York WS, Ye Z-H (2005) Arabidopsis fragile Fiber8, which encodes a putative glucuronyltransferase, is essential for normal secondary wall synthesis. Plant Cell 17: 3390–3408

**Supporting Information**

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

**Table S1.** Mutants used in the screen. All were confirmed to be homozygote. Col-0, *Arabidopsis thaliana* ecotype Columbia; Ws-o, *Arabidopsis thaliana* ecotype Wassilewskija.