Analytical Py-GC/MS of Genetically Modified Poplar for the Increased Production of Bio-aromatics

Gorugantu SriBala a, Hilal Ezgi Toraman a, Steffen Symoens a, Annabelle Déjardin b, Gilles Pilate b, Wout Boerjan c,d, Frederik Ronse e, Kevin M. Van Geem k, Guy B. Marin a

a Ghent University, Laboratory for Chemical Technology, Technologiepark 125, 9052 Ghent, Belgium
b Institut National de la Recherche Agronomique (INRA), Unité de Recherche 0588, Amélioration, Génétique et Physiologie Forestières, 45075 Orléans, France
c Ghent University, Department of Plant Biotechnology and Bioinformatics, Technologiepark 71, 9052 Ghent, Belgium.
d VIB Center for Plant Systems Biology, Technologiepark 71, 9052 Ghent, Belgium.
e Ghent University, Department of Biosystems Engineering, Coupure Links 653, 9000 Ghent, Belgium.

A R T I C L E  I N F O
Article history:
Received 20 December 2018
Received in revised form 11 April 2019
Accepted 12 April 2019
Available online 25 April 2019

Keywords:
Genetically modified poplar
Principal component analysis
Analytical fast pyrolysis
Lignin
Phenolic compounds

A B S T R A C T
Genetic engineering is a powerful tool to steer bio-oil composition towards the production of specialty chemicals such as guaiacols, syringols, phenols, and vanillin through well-defined biomass feedstocks. Our previous work demonstrated the effects of lignin biosynthesis gene modification on the pyrolysis vapour compositions obtained from wood derived from greenhouse-grown poplars. In this study, field-grown poplars downregulated in the genes encoding CINNAMYL ALCOHOL DEHYDROGENASE (CAD), CAFFEIC ACID O-METHYLTRANSFERASE (COMT) and CAFFEYL-CoA O-METHYLTRANSFERASE (CCoAOMT), and their corresponding wild type were pyrolysed in a Py-GC/MS. This work aims at capturing the effects of downregulation of the three enzymes on bio-oil composition using principal component analysis (PCA). 3,5-methoxytoluene, vanillin, coniferyl alcohol, 4-vinyl guaiacol, syringol, syringaldehyde, and guaiacol are the determining factors in the PCA analysis that are substantially affected by COMT, CAD and CCoAOMT enzyme downregulation. COMT and CAD downregulated transgenic lines proved to be statistically different from the wild type because of a substantial difference in S and G lignin units. The scAD line lead to a significant drop (nearly 51%) in S-lignin derived compounds, while CCoAOMT downregulation affected the least (7–11%). Further, removal of extractives via pretreatment enhanced the statistical differences among the CAD transgenic lines and its wild type. On the other hand, COMT downregulation caused 2-fold reduction in S-derived compounds compared to G-derived compounds. This study manifests the applicability of PCA analysis in tracking the biological changes in biomass (poplar in this case) and their effects on pyrolysis-oil compositions.

© 2019 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction
Biomass fast pyrolysis has gained enormous attention because of its potential to generate large amounts of bio-oil, which is an alternative to liquid fossil fuels. A detailed analysis of bio-oil suggests the presence of a variety of organic acids, aldehydes, ketones, phenols, guaiacols, and syringols, which are considered to be specialty chemicals [1–3]. The composition of bio-oil depends primarily on the feedstock and process parameters such as temperature, residence time, biomass particle sizes and reactor configuration [2–7]. A thorough understanding of the process parameters and biomass composition influencing fast pyrolysis product distribution is needed to optimize the process for large-scale applications.

So far, the focus of the research has been on understanding pyrolysis kinetics and achieving high bio-oil yields through unique catalytic or reactor designs [7]. Catalytic applications such as hydro-deoxygenation have proven to improve the usability of bio-oils by reducing its oxygen content [8–10]. Process intensification studies suggest an increase in bio-oil yield by improving heat and mass transfer between biomass and inert gases [11]. In the recent past, attempts have been made to alter the composition of bio-oil through well-defined feedstocks. The composition of biomass is altered with the help of genetic modification in plant species by down-regulating specific genes encoding the
enzymes of phenylpropanoid and monolignol biosynthetic pathways [12,13]. As a consequence, the modified feedstock composition could influence the product distribution in fast pyrolysis. In this way the reactions could be selectively directed towards an increased production of high-value chemicals, improving the profitability and industrial relevance of fast pyrolysis of biomass [14]. However, full-fledged analysis of the impact of genetic engineering on biomass fast pyrolysis has yet to be performed.

Lignin, the primary source of phenolic compounds in pyrolysis oil, is synthesized from monolignols via bond linkages such as β-O-4, β-5, β-3, yielding p-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) units derived from p-coumaryl, coniferyl and sinapyl alcohols, respectively [15]. While these linkages between lignin monomers are responsible for its linear structure, 4-O-5 and 5–5 couplings were assumed to yield a branched structure of lignin. However, according to a recent work of Ralph et al. [16] lignin structure is argued to be mostly linear due to lack of structural evidence for etherified branching in lignin chains [16]. The biosynthesis pathway of lignin involves a wide range of enzymatic reactions (as shown in Fig. 1). Lignin composition varies from one plant species to another, even more so in different cell types and cell wall layers [17]. CAFFEYL-CoA O-METHYLTRANSFERASE (CCoAOMT) converts caffeoyl-CoA to feruloyl-CoA, which is a precursor to coniferaldehyde and sinapaldehyde. The enzyme, CAFFEIC ACID O-METHYLTRANSFERASE (COMT) is responsible for the formation of sinapaldehyde, which is the precursor of sinapyl alcohol. CINNAMYL ALCOHOL DEHYDROGENASE (CAD) is involved in the formation of all the three monolignols [12,17]. Re-directing these pathways by downregulating one or more of the corresponding genes via genetic modification to yield tailor-made lignin could enable control over the pyrolysis product distribution [14,18].

Various poplar (Populus tremula x alba) trees which are genetically modified for lignin amount and composition have been made [18,19]. Lignin from CCoAOMT downregulated trees showed an increase of 11% in the S/G ratio [20]. On the other hand, COMT downregulation lead to a reduction in the S/G ratio by approximately 50% accompanied by the appearance of 5-hydroxyguaiaacyl units [21,22]. The CAD suppressed transgenic lines showed an increased production of benzaldehyde, vanillin, syringaldehyde and extractable lignin content due to the incorporation of cinnamaldehydes into the lignin structure [19,22–24]. However, no significant change was observed in the S/G [23]. Pyrolysis of wood derived from COMT downregulated greenhouse-grown Populus tremula x alba revealed a three-fold decrease in the pyrolysis products corresponding to syringyl (S) units while CCoAOMT down-regulation lead to a decline in G-derived products by 1.6 times, as reported by Toraman et al. [14]. There have been very few studies on fast pyrolysis of using genetically engineered biomass. All the trees analyzed so far were grown either in greenhouse or in controlled environments [14,18]. Rencoret et al. [25] utilized fast pyrolysis as a quick technique to determine the effect of gene modification in greenhouse grown poplar [25]. They reported that overexpression of FSH in poplar lead to a substantial increase in S units. Pilate et al. [19] reported that the structural changes in lignin in field grown transgenic poplars were similar to those in greenhouse grown plants. However, in the case of COMT downregulated lines, the lignin structural changes were milder due to less suppressed enzymatic activity compared to that of greenhouse grown trees [19].

The current study focuses on deciphering the extent of COMT, CCoAOMT and CAD gene suppression, and the effect of pre-treatment on the wood derived from field-grown wild-type and transgenic poplar [17]. To the best of our knowledge, this is the first ever study on pyrolysis of genetically modified poplars downregulated in CAD genes. The primary aim of this study is to understand the effects of COMT, CCoAOMT and CAD downregulation on the relative abundance of pyrolysis products. The presence of extractable oligo and monophenols may interfere in the Py-GC-MS studies, and therefore, the samples were also pretreated to remove extractable phenolic compounds before the pyrolysis experiments. Principal component analysis (PCA) was utilized to analyze the extensive multi-dimensional experimental data set, which otherwise would be difficult to interpret. K-means clustering based on the Mahalanobis distance has been applied onto the score plots of PCA to investigate the statistical independence of the pyrolysis products obtained from the genetically modified and wild-type poplars. With the help of PCA and k-means clustering, the effect of the downregulation of COMT, CCoAOMT and CAD on the yields of S and G derived products is understood.

2. Materials and Methods

2.1. Lignocellulosic Biomass Samples

Genetically modified poplar trees and the corresponding wild types were grown in a field in Orleans, France. Three types of transgenic lines were produced with COMT, CAD, and CCoAOMT downregulation as described in Van Doorsselaere et al. [21], Baucher et al. [23], Lapierre

![Fig. 1. Lignin biosynthesis pathways leading to the formation of H, G, and S units. PAL, PHENYLALANINE AMMONIA-LYASE; C4H, CINNAMATE 4-HYDROXYLASE; 4CL, 4-COUMARATE:CoA LIGASE; CH3, p-COUMARATE 3-HYDROXYLASE; CSE, CAFFEYL SHIKIMATE-ESTERASE; HCT, p-HYDROXYCINNAMOYL-CoA:QUINATE/SHIKIMATE; CCoAOMT, CAFFEYL-CoA O-METHYLTRANSFERASE; CCR, CINNAMOYL-CoA REDUCTASE; F5H, FERULATE 5-HYDROXYLASE; COMT, CAFFEIC ACID O-METHYLTRANSFERASE; and CAD, CINNAMYL ALCOHOL DEHYDROGENASE [15,17].](image-url)
et al. [22] and Meyermans et al. [20]. The COMT (Li09, Li11) and CAD lines (Li18, Li21, Li22) and corresponding control line (Tin) were grown in a single field that was planted in June 2008, coppiced in March 2010 and harvested after 2 years of growth in February 2012. CCoAOMT down-regulated lines were grown on a neighbouring field along with their corresponding wild-type (reference Tbr). On this second plot, trees were planted in May 2009, coppiced in 2010 at the same time as the other plot, and collected in February 2012 as well. Both fields were divided into five blocks, with each block containing 24 clonal replicates per line (Fig. 2). Poplar was first debarked, and the

Fig. 2. Plan of field trial at Orleans, corresponding to the sample list provided in Table 1. CCoAOMT downregulated poplar trees along with their wild-type, Tbr, were grown in a separate field compared to all other poplar trees. Each of the poplar species have at least 5 repeats as shown in the figure, which are represented as block numbers in the Table 1.
dried wood chips collected from each line and each block were first ground using a 6 mm grid and then further ground using a high-frequency grinder (Retch 300, 200 Hz) and sieved to remove particles larger than 500 μm. Table 1 lists the samples used in the current work. The composition of lignin in terms of G or S units was obtained using the thioacidolysis procedure developed by Lapierre et al. [26,27]. The procedure involves removal of cell-wall extractives by incubating small amounts of sample (~12 mg) in water at 98 °C, followed by ethanol at 76 °C, chloroform at 59 °C and acetone at 54 °C. Each of these incubation steps was performed for 30 min, and the supernatant was removed by centrifuging at 14000 rpm for 3 min. The dried sample pellets free from cell-wall are then treated with a reaction mixture of boron trifluoride, ethanediol and dioxane. The sample vials were filled with liquid N₂ vapours to maintain the inert atmosphere and placed in a water bath at 98 °C for 4 h. Further, tetracosane was added as an internal standard to each of the vials before extracting with dichloromethane and water. The organic phase was then pipetted into a new Eppendorf and dried using speedvac. S and G unit representatives present in the dried sample were first dissolved in dichloromethane and then derivatised using N, O-bis(trimethylsilyl)acetamide for GC–MS analysis.

2.2. Pre-treatment

Transgenic samples from block five were pre-treated to remove extractives by incubating them in various solvents, separately [28,29]. The solvents were chosen such that the structure of cellulose, hemicellulose, and lignin remained intact during the pre-treatment [28,29]. Depending on the type of solvent, the incubation temperature was set, as shown in Table 2. About 12 mg of each biomass sample was incubated in 1 ml of each solvent at 750 rpm. After that, the supernatants were removed by centrifugation at 14000 rpm. Samples were then air dried to evaporate the remaining solvents. The incubation time for all the solvent pretreatments was 30 min.

2.3. Fast Pyrolysis Experiments

The transgenic and wild-type poplar samples were pyrolysed in a multi-shot pyrolyser (EGA/PY-3030D, Frontier Laboratories, Japan). The furnace was calibrated to read the centerline temperature of the quartz reactor. A deactivated stainless steel sample cup (Eco-cup SF, Frontier Laboratories, Japan) was loaded with 0.3 to 0.4 mg of fine biomass powder and was dropped into the pre-heated reactor at 500 °C, ensuring rapid pyrolysis. Pyrolysis vapours were directly swept into the injector port (300 °C) of the GC (Thermo Trace GC Ultra) using helium as carrier gas set at a constant flow rate of 100 ml/min with an injector split ratio of 1:100. Pyrolysis products were separated using a Restek RTX-1701 column (60 m x 0.25 mm, 0.25 μm) connected to a mass spectrometer (ISQ-MS) with a scan rate set from 25 to 350 a.m.u. The GC oven was held for 3 min at an initial temperature of 40 °C and was subsequently heated to 280 °C at 5 °C/min. The oven was then held at 280 °C for 1 min. After each injection, a blank analysis was carried out to verify potential residual bleeding of the column and leftovers from the previous experiments. Peak integration and alignment were performed in Xcalibur software using a baseline window of 200, area to noise factor of 100 and a peak noise factor value of 10. The products were identified using the NIST library and quantified by normalising the ion current peak areas with the total area of all the compounds. The methodology of the process is pictorially represented in Fig. 3.

Table 1

| Name   | Published name of poplar line | Blocks | G (μg) | S (μg) | S/G ratio  
|--------|-------------------------------|--------|--------|--------|-----------
| Tin    | wild-type                     | 1–5    | 63 ± 13| 161 ± 20| 2.63   
| Thb    | wild-type for CCoAOMT         | 1–5    | 56 ± 10| 150 ± 28| 2.65   
| Li02   | asCCoAOMT101                 | 1–5    | 52 ± 10 | 149 ± 22 | 2.84  
| Li03   | sCCoAOMT416                  | 1–5    | 54 ± 6  | 147 ± 10 | 2.66  
| Li04   | sCCoAOMT429                  | 1–5    | 56 ± 8  | 149 ± 16 | 2.66  
| Li09   | asCOMT28                     | 1–5    | 58 ± 14 | 133 ± 30 | 2.21  
| Li11   | asCOMT10B                    | 1–5    | 62 ± 9  | 92 ± 12  | 1.34  
| Li18   | asCAD52                     | 1–5    | 46 ± 16 | 124 ± 20 | 2.65  
| Li21   | asCAD21                     | 1–5    | 48 ± 16 | 131 ± 30 | 2.72  
| Li22   | scCAD1                      | 3–5    | 41 ± 20 | 114 ± 20 | 2.71  

Table 2

<table>
<thead>
<tr>
<th>Solvent</th>
<th>Incubation temperature (°C)</th>
<th>Incubation time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>98</td>
<td>30</td>
</tr>
<tr>
<td>Ethanol</td>
<td>76</td>
<td>30</td>
</tr>
<tr>
<td>Chloroform</td>
<td>59</td>
<td>30</td>
</tr>
<tr>
<td>Acetone</td>
<td>54</td>
<td>30</td>
</tr>
</tbody>
</table>

3. Principal Component Analysis

Principal component analysis (PCA) of the quantified peak areas was performed using MATLAB (R2014b) software. The methodology is similar to the one described by Toraman et al. [14]. Briefly, PCA analysis provides a statistical analysis of the variation of the selectivity differences between transgenic and control lines, and hence, indicating the effect of genetic modification on the pyrolysis vapours. The first principal component accounts for as much of the variability in the data as possible (PC1), and each succeeding element accounts for the remaining variability (PC2, PC3, etc.). This procedure was repeated until ideally the total variance obtained in the original data set was explained and the resulting PCs formed a new basis [30–32]. Variations in transgenic and wild-type biomass and their pyrolysis products were studied with the help of score and loading plots, as explained by Toraman et al. [14].

The K-means clustering algorithm was applied to identify groups among wild-type and transgenic poplar samples based on the PCA [33–35]. The method has been described elsewhere in detail [14]. Briefly, K-means clustering results in partitioning the data space into some regions, called Voronoi cells. In this work, the Mahalanobis distance (MD) method was used to formulate the clusters [32,36–39]. Locus of the ellipse defining a group of data points which are similar to the centroid was calculated using eq. 1 with an assumed level of certainty [14].

\[ T^2 = \frac{p}{n-p} \left[ \frac{(n-1)}{p} F_{p,n-p,1-\alpha} = MD^2 \right. \] (1)

where p is the number of principal components, n is the number of observations, F is the F-statistic value, and 1-α is the confidence interval percentage.

4. Results and Discussion

Transgenic and wild-type poplar samples were subjected to fast pyrolysis to examine the effect of down-regulation of a specific gene in the lignin biosynthetic pathway on the release of products originating from lignin. At first, the pyrolysis vapour composition of non-pretreated transgenic samples was compared to that of their respective non-pretreated wild-type samples. Further, the transgenic lines were compared with their pre-treated counterparts to understand the influence of the extractives on the product spectrum. All the samples were analyzed at least in triplicate. In total 46 compounds were identified in all pyrograms, including the products originating from holocellulose (i.e., cellulose and hemicellulose (C), syringyl lignin (L-S), guaiacyl lignin (L-G), p-hydroxyphenyl lignin (L-H) or lignin in general (L)). All the detected compounds along with their retention time (on the
presented GC configuration) and origin are listed in Table 3. Primarily, L-G and L-S derived pyrolysis compounds produced from the transgenic poplar were relatively lower than their corresponding wild type. A functional group-based analysis of lignin-derived pyrolysis products in terms of organics containing methoxyphenols, dimethoxyphenols, and phenols is presented in Fig. 4. Overall, downregulation of COMT,

![Fig. 3. Methodology of genetically modified biomass pyrolysis, and subsequent PCA analysis of the product selectivities.](image)

<table>
<thead>
<tr>
<th>R.T (min)</th>
<th>Compound</th>
<th>Origin</th>
<th>Tbr</th>
<th>Li02</th>
<th>Li03</th>
<th>Li04</th>
<th>Tin</th>
<th>Li09</th>
<th>Li11</th>
<th>Li18</th>
<th>Li21</th>
<th>Li22</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.38</td>
<td>Carbon dioxide</td>
<td>C1</td>
<td>7.63</td>
<td>1.17</td>
<td>1.42</td>
<td>1.23</td>
<td>7.38</td>
<td>1.20</td>
<td>1.21</td>
<td>1.27</td>
<td>1.33</td>
<td>1.45</td>
</tr>
<tr>
<td>4.51</td>
<td>Ethane</td>
<td>C2</td>
<td>3.48</td>
<td>1.20</td>
<td>1.38</td>
<td>1.33</td>
<td>3.78</td>
<td>1.16</td>
<td>1.21</td>
<td>1.17</td>
<td>1.21</td>
<td>1.25</td>
</tr>
<tr>
<td>4.87</td>
<td>Acetaldehyde</td>
<td>C3</td>
<td>1.17</td>
<td>1.13</td>
<td>1.24</td>
<td>1.36</td>
<td>1.07</td>
<td>0.93</td>
<td>0.87</td>
<td>0.79</td>
<td>0.81</td>
<td>0.83</td>
</tr>
<tr>
<td>5.01</td>
<td>Furan</td>
<td>C4</td>
<td>0.19</td>
<td>0.11</td>
<td>0.17</td>
<td>0.22</td>
<td>0.16</td>
<td>0.21</td>
<td>0.25</td>
<td>0.28</td>
<td>0.32</td>
<td>0.35</td>
</tr>
<tr>
<td>5.92</td>
<td>2-Propenal (propanal-2-one)</td>
<td>C5</td>
<td>0.37</td>
<td>0.12</td>
<td>0.27</td>
<td>0.30</td>
<td>0.38</td>
<td>0.34</td>
<td>0.32</td>
<td>0.33</td>
<td>0.36</td>
<td>0.41</td>
</tr>
<tr>
<td>6.05</td>
<td>Acetic anhydride</td>
<td>C6</td>
<td>3.48</td>
<td>1.13</td>
<td>1.10</td>
<td>1.16</td>
<td>1.17</td>
<td>1.18</td>
<td>1.19</td>
<td>1.20</td>
<td>1.21</td>
<td>1.22</td>
</tr>
<tr>
<td>6.51</td>
<td>Furan, 2-methyl-</td>
<td>C7</td>
<td>0.19</td>
<td>0.17</td>
<td>0.22</td>
<td>0.29</td>
<td>0.31</td>
<td>0.35</td>
<td>0.38</td>
<td>0.41</td>
<td>0.44</td>
<td>0.47</td>
</tr>
<tr>
<td>7.72</td>
<td>2,3-butanedione</td>
<td>C8</td>
<td>0.98</td>
<td>1.10</td>
<td>1.32</td>
<td>1.24</td>
<td>0.98</td>
<td>1.17</td>
<td>1.24</td>
<td>1.31</td>
<td>1.38</td>
<td>1.45</td>
</tr>
<tr>
<td>8.14</td>
<td>Propanoic acid, anhydride</td>
<td>C9</td>
<td>0.60</td>
<td>1.07</td>
<td>1.13</td>
<td>1.24</td>
<td>0.98</td>
<td>1.17</td>
<td>1.24</td>
<td>1.31</td>
<td>1.38</td>
<td>1.45</td>
</tr>
<tr>
<td>8.96</td>
<td>Hydroxyacetaldehyde</td>
<td>C10</td>
<td>0.69</td>
<td>1.04</td>
<td>0.96</td>
<td>0.97</td>
<td>1.15</td>
<td>1.01</td>
<td>1.07</td>
<td>1.12</td>
<td>1.21</td>
<td>1.30</td>
</tr>
<tr>
<td>10</td>
<td>Acetic acid</td>
<td>C11</td>
<td>12.64</td>
<td>1.03</td>
<td>0.96</td>
<td>0.97</td>
<td>12.41</td>
<td>1.01</td>
<td>1.07</td>
<td>1.12</td>
<td>1.21</td>
<td>1.30</td>
</tr>
<tr>
<td>11.34</td>
<td>2-Propanol, 1-hydroxy-</td>
<td>C12</td>
<td>4.21</td>
<td>0.99</td>
<td>1.06</td>
<td>1.08</td>
<td>3.98</td>
<td>1.05</td>
<td>1.09</td>
<td>1.12</td>
<td>1.08</td>
<td>1.07</td>
</tr>
<tr>
<td>12.93</td>
<td>Furan, 3-methyl-</td>
<td>C13</td>
<td>1.67</td>
<td>1.13</td>
<td>1.28</td>
<td>1.36</td>
<td>1.01</td>
<td>1.17</td>
<td>1.24</td>
<td>1.31</td>
<td>1.38</td>
<td>1.45</td>
</tr>
<tr>
<td>14.37</td>
<td>1,2-Ethanediol</td>
<td>C14</td>
<td>0.98</td>
<td>1.10</td>
<td>1.32</td>
<td>1.24</td>
<td>0.98</td>
<td>1.17</td>
<td>1.24</td>
<td>1.31</td>
<td>1.38</td>
<td>1.45</td>
</tr>
<tr>
<td>16.85</td>
<td>Furfural</td>
<td>C15</td>
<td>3.80</td>
<td>1.02</td>
<td>0.83</td>
<td>0.72</td>
<td>3.61</td>
<td>1.07</td>
<td>1.12</td>
<td>1.07</td>
<td>1.12</td>
<td>1.17</td>
</tr>
<tr>
<td>18.19</td>
<td>2-Furanmethanol</td>
<td>C16</td>
<td>0.82</td>
<td>0.86</td>
<td>0.61</td>
<td>0.75</td>
<td>0.59</td>
<td>0.78</td>
<td>0.80</td>
<td>0.70</td>
<td>0.76</td>
<td>0.80</td>
</tr>
<tr>
<td>18.28</td>
<td>2(3H)-Furanone, dihydro-4-hydroxy-</td>
<td>C17</td>
<td>0.38</td>
<td>0.99</td>
<td>1.02</td>
<td>0.93</td>
<td>0.94</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>25.07</td>
<td>β-D-Glucopyranose, 1,6-anhydro- (levoglucosan)</td>
<td>C30</td>
<td>6.73</td>
<td>0.88</td>
<td>1.18</td>
<td>1.36</td>
<td>1.43</td>
<td>1.37</td>
<td>1.96</td>
<td>1.91</td>
<td>1.74</td>
<td>1.77</td>
</tr>
<tr>
<td>25.86</td>
<td>Phenol, 2-methyl-</td>
<td>C18</td>
<td>2.37</td>
<td>0.92</td>
<td>0.55</td>
<td>0.70</td>
<td>2.44</td>
<td>0.68</td>
<td>0.74</td>
<td>0.60</td>
<td>0.61</td>
<td>0.52</td>
</tr>
<tr>
<td>29.34</td>
<td>3,5-Dimethoxytoluene</td>
<td>C19</td>
<td>0.11</td>
<td>0.77</td>
<td>1.10</td>
<td>1.46</td>
<td>0.07</td>
<td>2.18</td>
<td>1.73</td>
<td>1.97</td>
<td>1.61</td>
<td>1.71</td>
</tr>
<tr>
<td>37.28</td>
<td>Coniferyl alcohol (trans)</td>
<td>C20</td>
<td>1.67</td>
<td>0.93</td>
<td>0.86</td>
<td>0.80</td>
<td>1.64</td>
<td>0.93</td>
<td>0.86</td>
<td>0.80</td>
<td>0.83</td>
<td>0.80</td>
</tr>
<tr>
<td>40.09</td>
<td>Phenol</td>
<td>C21</td>
<td>1.67</td>
<td>0.93</td>
<td>0.86</td>
<td>0.80</td>
<td>1.64</td>
<td>0.93</td>
<td>0.86</td>
<td>0.80</td>
<td>0.83</td>
<td>0.80</td>
</tr>
<tr>
<td>42.58</td>
<td>Syringaldehyde</td>
<td>C22</td>
<td>1.67</td>
<td>0.93</td>
<td>0.86</td>
<td>0.80</td>
<td>1.64</td>
<td>0.93</td>
<td>0.86</td>
<td>0.80</td>
<td>0.83</td>
<td>0.80</td>
</tr>
<tr>
<td>45.81</td>
<td>Acetosyringone</td>
<td>C23</td>
<td>1.67</td>
<td>0.93</td>
<td>0.86</td>
<td>0.80</td>
<td>1.64</td>
<td>0.93</td>
<td>0.86</td>
<td>0.80</td>
<td>0.83</td>
<td>0.80</td>
</tr>
<tr>
<td>49.61</td>
<td>Sinapaldehyde</td>
<td>C24</td>
<td>1.67</td>
<td>0.93</td>
<td>0.86</td>
<td>0.80</td>
<td>1.64</td>
<td>0.93</td>
<td>0.86</td>
<td>0.80</td>
<td>0.83</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Table 3
List of the main identified compounds of the micro-pyrolysis experiments, average peak surface area percentages of Tin and Tbr wild-type poplars and ratios of average peak surface area% of transgenic lines relative to average peak surface area% of their respective control lines.
CCoAOMT and CAD led to a drop in methoxy phenols and dimethoxy phenols up to a factor of 1.6 (Lii22 and Lii11, respectively). Simple phenolic products, however, tended to increase up to 1.6 times during the transgenic poplar (Lii22) pyrolysis.

To extract a correlation in the data set six PCA models were developed, corresponding to each of the poplar types. At first, the number of principal components that need to be retained for each kind of transgenic and wild-type poplar pair were determined with the help of scree plots. The scree plot shows the relative contribution of the principal components to the total variance in the data set for either non-pretreated or pretreated samples. Principal components are included until the first three principal components represented 57.67%, 26.66% and 6.53% of the total variance. The scree plots of untreated and pretreated samples are shown in Fig. 5a and Fig. S1, respectively. The score and loading plots for the principal components PC1 vs PC2 and PC1 vs PC3 of untreated samples are shown in Fig. 5c and d.

The score plot of PC1 vs PC2 (5b) shows two clusters of COMT and Tin samples with clustering performed at a confidence interval of 85%. Based on the corresponding loading plot (5c), it can be observed that the most considerable positive contributions to the first principal component (PC1) come from G and S lignin units, especially syringol (L-S1), 4-allylsyringol (L-S2), syringaldehyde (L-S3), acetosyringone (L-S4), sinapaldehyde (L-S5), 4-ethylguaiacol (L-G3), 4-vinyl guaiacol (L-G4), 4-propyl guaiacol (L-G6), and coniferyl alcohol (L-G7). The other positive contributions to PC1 come from 2-methyl phenol (L-H2), 5-hydroxymethylfurfural (C28), 2-furanmethanol (C21), hydroxycetaldehyde (C11), furfural (C20), methanol (C4), 1,2-ethanediol monoacetate (C17), and vanillin (L-G5).

4.1.1. Comparison of COMT Lines to Tin Wild-Type

In the case of the COMT model, one control line (Tin wild type) and two transgenic COMT lines (Lio9 and Lii11) were considered. PC1 and PC2 described 53.37% and 27.89% of the total variance in the scree plot for untreated COMT and Tin control samples. On the other hand, the pre-treated COMT and Tin samples, the first three principal components represented 56.67%, 26.66%, and 6.53% of the total variance. The scree plots of untreated and pretreated samples are shown in Fig. 5a and Fig. S1, respectively. The score and loading plots for the principal components PC1 vs PC2 and PC1 vs PC3 of untreated samples are shown in Fig. 5c and d.

The loadings of the pretreated samples (Fig. S1), on the other hand, indicate that all the lignin-derived products (except L-H1 and L-2) contribute positively to PC1. Moreover, pre-treated COMT lines are found in the left half-plane of the score plot of PC1 vs PC2, i.e., on the negative axis of PC1. The wild-type is pulled towards PC1 = 0 by the deviating sample 3, although sample 1 and 2 are clearly in the positive part of the PC1 axis. The result for the pretreated COMT samples is similar.
that for the non-pretreated samples, except that L-G5 is not zero for PC1 and is less pronounced for its contribution to PC2. Hence, it is concluded that the pre-treatment of the COMT samples does not have a significant influence on the pyrolysis vapour composition of the genetically engineered poplar and wild type samples.

### 4.1.2. Comparison of CAD Lines to Tin Wild-Type

The CAD model consists of Tin wild type and three transgenic lines, viz. Li18, Li21, and Li22. First, three principal components of the non-pretreated samples accounting for 46.93%, 22.60% and 14.03% of the total variance. These values remain similar for the three components of pretreated CAD and Tin samples with 54.21%, 23.06%, and 9.67%, respectively. These are represented by scree plots in Fig. 6a and 7a.

In this model, the untreated poplar lines are grouped in three clusters, viz. Tin wild-type, Li22 CAD, and Li18 and Li21 CAD lines together in a third cluster, with a confidence interval of 85%. The three clusters are separated mainly by the first principal component. Interestingly, the transgenic line Li22 is on the positive axis of PC1 while the wild-type Tin is on the negative axis. The third cluster of Li18 and Li21 is at PC1 = 0 implying that no mutual differentiation can be drawn for these two types of samples. At this juncture, it is interesting to note the clear discrimination between Li18 and Li21, and Li22. Li18 and L21 are lines in which the CAD gene was downregulated using an antisense approach, whereas to generate Li22, the downregulation of CAD was achieved through the introduction of a sense construct. This suggests that the two strategies for genetic modification result in a different material, presumably because of the different levels of CAD downregulation. The substantial positive contribution from lignin-derived products to the first and second principal components comes from 3,5-dimethoxytoluene (L-2), 4-propylguaiacol (L-G6), vanillin (L-G5), 4-ethylguaiacol (L-G3) and 2-methylphenol (L-H2). These products are found in the highest selectivities for the Li22 CAD line and the lowest selectivities for the Tin wild type (score plot, Fig. 6b). Coniferyl alcohol (L-G7), 4-vinyl guaiacol (L-G4), syringol (L-S1), syringaldehyde (L-S3) and guaiacol (L-G1) provide a significant negative contribution to the first principal component and are found in the lowest selectivities for Li22 CAD and the highest selectivities for the Tin control line. For the second principal component, coniferyl alcohol and 4-vinyl guaiacol contribute the most in the positive direction, while phenol and 4-methyguaiacol provide a substantial negative contribution. S-units like L-S3, L-S4, and L-S5 and L-H1 have positive contributions to PC3 while L-G1, L-G7, and L-S1 are closer to PC3 = 0. The remaining lignin products are on the negative axis of the third principal component.
Based on Fig. 6b and c it could be concluded that the Li22 CAD line produces higher amounts of coniferyl alcohol and 4-vinyl guaiacol and lower amounts of phenol and 4-methylguaiacol during fast pyrolysis, compared to the Li18 and Li21 CAD lines. Since the lignin-derived products are not grouped but spread over all quadrants as can be seen in the loading plots Fig. 6c & d, it is difficult to conclude the effect of CAD down-regulation on the change of the amount of G, S and H units in poplar lignin based on the pyrolysis results. Thioacidolysis data (Table 1) indicates a reduction in both S and G representative units in Li18, Li21 and Li22 transgenic lines. However, the ratio of S/G seems to either remain similar or increase compared to Tin. Downregulation of CAD decreased the relative amount of G-lignin derived products about 25% based on the peak areas. The aCAD downregulation leads to a decrease of 21% in S-lignin derived compounds while sCAD downregulation showed as high as 51% decrease. This result is in line with the thioacidolysis data and the biosynthesis pathway.

As no distinct clusters could be observed in the score plot of PC1 vs PC2 for the pretreated samples, the PC1 vs PC3 score plots was considered for clustering purposes (Fig. 7). The loading plots for PC1, PC2, and PC3 of pretreated samples are shown in Fig. 6c and d. All originating lignin products except phenol (L-H1) and 3,5-dimethoxytoluene (L-2) contribute positively to PC1. PC3 contains stronger positive contributions from S-type products, especially syringaldehyde (L-S3) and sinapaldehyde (L-S5). This is very interesting to observe as sinapaldehyde accumulation is very high in the lignin of CAD deficient plants, but negligible amounts of coniferyl aldehyde (L-G7). The strongest negative contributions come from 4-allylsyringol (L-S2) and G-type products. In the score plot (Fig. 7b) K-means clustering (p < .15) with the Mahalanobis distance method results in 4 sample groups, one for each wild-type and transgenic lines. All three CAD lines are located in the half-plane for negative values of PC1 which indicates that the total amount of S and G lignin-derived products is lower in the pretreated CAD down-regulated samples compared to the Tin control line. For the non-pretreated samples, this was not observed. Therefore, for the CAD lines, the differences in all the three genetically modified materials were visible only after removal of extractives. This observation is substantiated by the work of Van Acker et al. [24]. According to their study CAD deficient plants accumulate massive amounts of sinapaldehyde (5(S–8)S) dimeric species as soluble phenolics. Pre-treatment of CAD deficient poplar samples with various solvents

---

**Fig. 6.** Comparison of untreated CAD lines and Tin wild type. (a) Scree plot showing the relative weight percentages of each principal component. (b) Score plot of PC1 vs PC2, with three clusters of wild-type, Li22 and the third with Li18 and Li20 transgenic poplars. The lines are clustered with a significance level of 85%. (c) and (d) Loading plots of PC1 vs PC2 and PC1 vs PC3, indicating the data points that have the highest and lowest contributions to the principal components. (Color codes are as follows: Pink, Cellulose; Red, Lignin-H; Cyan, Lignin-S; Blue, Lignin-G; and Yellow, generic Lignin compounds, i.e., L-1 and L-2).
mentioned in Section 2.2 results in the extraction of the soluble pheno-
olics from the plant material.

4.1.3. Comparison of CCoAOMT Down-Regulation to Tbr Wild-Type

The dataset for the CCoAOMT model consists of Tbr wild-type and
three CCoAOMT down-regulated lines (Li02, Li03, and Li04). The scree
plot in Supporting Information, corresponding to non-pretreated
CCoAOMT and Tbr samples, indicates the proportion of variance de-
described by the first three principal components as 47.91%, 14.19% and
12.78% of the total variance. While for the case of pretreated CCoAOMT
and Tbr control samples, the first three principal components account
for 30.36%, 25.48% and 16.54% of the total variance (Fig. 8a).

From the score plot (Fig. S2) between principal components PC1 and
PC2 three clusters were formed, one corresponding to Tbr control line
and Li02 together and the remaining two correspond to the data points
of Li03 and Li04 (p < .20). Further, separation in the loading plot was ob-
served primarily along the first principal component, PC1. The data of
Li02 samples have been clustered along with the control line, and no
meaningful conclusions could be drawn.

As the data points of Li03 and Li04 could not be decoupled from each
other, a new data set was considered for PCA analysis. In this case, only
lignin products have been studied from the pyrolysis of Tbr and
CCoAOMT downregulated samples. The data point number 10 (Li03)
was removed from PCA analysis as it was an outlier. Variance values of
the first three principal components for the new data set are 53%,
21.8%, and 11.7%, respectively (8a). Three clusters could be identified
with a confidence interval of 95%, namely Li02, Tbr and a combination
of Li03 and Li04 as seen in Fig. 8b. The clusters are shown in the score
plot, Fig. 8b. Surprisingly, the Li02 line (asCCoAOMT) differentiates
strongly on the PC1 from the other two transgenic lines, which are so-
called sense (sCCoAOMT) lines. The effect of CCoAOMT downregulation
is very evident in the new score plot. The loading plot between PC1
and PC2 also suggests that all the G-derived compounds contribute neg-
atively to PC1, while the S-derived compounds syringaldehyde (L-S3),
acetosyringone (L-S4) and sinapaldehyde (L-S5) contribute positively
to both PC1 and PC2 (Fig. 8c). According to the pathway shown in the
Fig. 1, CCoAOMT downregulation should reduce the formation of
coniferaldehyde and subsequently, sinapaldehyde. The score plots
could be indicative that the amount of G units in the Li03 and Li04

Fig. 7. Comparison of pre-treated CAD lines and Tin wild type. (a) Scree plot showing the relative weight percentages of each principal component. (b) Score plot of PC1 vs PC3, with four clusters of wild-type, Li22, Li18, and Li20 transgenic poplars. The lines are clustered with a significance level of 85%. (c) and (d) Loading plots of PC1 vs PC2 and PC1 vs PC3, indicating the data points that have the highest and lowest contributions to the principal components. (Color codes are as follows: Pink, Cellulose; Red, Lignin-H; Cyan, Lignin-S; Blue, Lignin-G; and Yellow, generic Lignin compounds, i.e., 1-1 and 1-2).
transgenic samples was substantially decreased upon the downregulation of CCoAOMT compared to the control line. The G and S units have been suppressed in Li03 and Li04, while there was a negative effect in Li02 samples. In accordance with PCA results, there was about 20–25% decrease in the G-lignin derived compounds upon the pyrolysis of all CCoAOMT poplar samples. However, only 7% decrease in S-derived compounds was observed during Li02 poplar pyrolysis. On the contrary, the greenhouse grown CCoAOMT downregulated lines showed an increase in S-derived products and a decrease in G-derived products relative to the control lines. Thioacidolysis data of field-grown poplar used in this work (Table 1) indicates a minimal reduction in S and G units due to genetic modification. The negative effect of Li02 observed in PCA studies could be attributed to higher S/G ratios than other two CCoAOMT downregulated lines Li03 and Li04.

No effects of extraction (pre-treatment of the biomass) on the fast pyrolysis selectivities of CCoAOMT downregulated samples could be concluded, as there was a random spread of data in the score plot (Fig. S3). It could be hypothesised that there was no accumulation of monophenolic extractives. One of two data points representing the Tbr wild-type (point no. 2) was grouped into the cluster of transgenic samples, and one data point from Li02 was consolidated with the wild-type samples, even at a confidence interval of 80%. This makes the transgenic cluster spread into the negative axis along PC1. The centroid of the cluster belonging to the transgenic poplar lines is on the positive axis of PC1, and the cluster representing wild-type samples is on the negative axis of PC1. Hence, there is a suppression of G and S units in the CCoAOMT lines, which is in line with the lower Klasson lignin described for these lines in literature [20]. The loading points of PC1 vs PC2 convey that almost all the guaiacyl and lignin-derived products contribute negatively to PC1 (Fig. S5). Only L-S3, L-G5, and L-G2 have positive projections on PC1. This could indicate that the quantity of G and S units in the CCoAOMT downregulated lines has been decreased only by a marginal amount as compared to the wild-type.

The effect of genetic engineering on large-scale pyrolysis is yet to be tested or on-going. However, with the insights obtained in the current work, it can be said that there is substantial potential to alter bio-oil compositions through engineered feedstock. Although, the modifications need to be more pronounced. Most recent CRISPR-Cas technology could facilitate more pronounced differences in biomass composition and thus in the resulting bio-oil [40,41]. This could help in generating...
wells with substantially improved compositions and better properties for fuel or chemical applications. Moreover, in combination with a catalyst, there is a possibility to achieve high selectivity towards speciality chemicals.

5. Conclusions

Micropyrolysis has proved to be an adequate analytical tool for studying fast pyrolysis of genetically modified poplar. In particular, the combination of micropyrolysis and PCA allows discriminating between the biomass feedstocks based on their lignin composition, in terms of G, H, and S units. About 46 compounds were identified in the pyrolysis vapours during Py-GC-MS studies performed using transgenic lines down-regulated for CCoAOMT, COMT, and CAD, and their control lines (Tin & Tbr). A functional group analysis of pyrolysis vapour product distributions suggested 2-fold drop in methoxy phenols and dimethoxy phenols due to COMT, CCoAOMT, CAD suppression in transgenic lines compared to their wild-type. According to the PCA analysis, COMT downregulated transgenic lines produced about 33% reduction in syringyl (S) derived compounds and a 15% reduction in guaiacyl (G) derived compounds. The result is in agreement with the biosynthetic pathway, wherein COMT suppression causes decreased incorporation of sinapyl alcohol and sinapaldehyde in the lignin polymer. sCAD downregulation leads to nearly 51% reduction in S lignin-derived compounds, while CCoAOMT downregulation caused only 7–11% reduction. Investigation of peak areas of pyrolysis vapours showed almost 20% decrease in C-derived products for all the transgenic lines. PCA applied in combination with K-means clustering was compared with thioacidolysis experiments and found that the models captured the trends in G and S composition reasonably well. The models could effectively discriminate between the wild-type and two CAD lines (L18 and L21), while the Li22 CAD line was separated into a third cluster, likely because of its different level of downregulation. CAD downregulation causes a strong accumulation of soluble phenolics and their extraction via pre-treatment made the effect of the gene modification on the pyrolysis products more pronounced. From this work, it is concluded that CAD downregulation in field-grown poplar is a promising strategy to steer the bio-oil composition towards relatively low amounts of G-lignin derived products. On the other hand, COMT downregulation is valuable to generate bio-oil with relatively more moderate quantities of G-and S-lignin compounds.

Acknowledgements

The authors thank Lennart Hoengenaart from VIB and Diana Vargas from LCT for their assistance in thioacidolysis experiments. The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement n° 290793 and the “Long-Term Structural Methusalem Funding by the Flemish Government” (IWT) is also acknowledged. The GBFOR experimental unit as well as Kevinader from Genobois platform (INRA Val de Loire) are warmly acknowledged for their involvement in the setting and maintenance of the GM poplar field trial, as well as the harvesting and pre-processing of the wood samples. This work has been funded by the FP7-KBBE-#211917 Energy/Poplar project.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2019.04.007.

References


