Supplement of

Microbial labilization and diversification of pyrogenic dissolved organic matter

Aleksandar I. Goranov et al.

Correspondence to: Patrick G. Hatcher (phatcher@odu.edu)

The copyright of individual parts of the supplement might differ from the article licence.
Table of Contents

Section 1. Chromophoric dissolved organic matter (CDOM) (Fig. S1) ................................................................. 3
Section 2. Presence-absence analysis of FT-ICR-MS formulas (Figs. S2-4).............................................................. 4
Section 3. H/C versus molecular weight plots (Figs. S5-8) .................................................................................... 7
Section 4. Evaluation of bio-resistant formulas (Fig. S9) ........................................................................................ 11
Section 5. Comparison of bio-produced formulas with estuarine and marine samples (Table S1) ......................... 12
Section 6. Analysis of variance (ANOVA) of bio-produced peptide-like formulas (Table S2) ............................ 13
Section 7. Oligopeptide sequences (Table S3) .................................................................................................... 14
Section 8. Kendrick mass defect plots of bio-produced formulas (Fig. S10-12)....................................................... 16
Section 9. Correlation analysis of molecular diversity and NMR data (Table S4) ................................................ 19
References ............................................................................................................................................................ 20
Section 1. Chromophoric dissolved organic matter (CDOM)

Figure S1. Total chromophoric dissolved organic matter (CDOM) content of pyDOM leachates before (blue) and after (green) 10-day biotic incubations. Total CDOM content is reported as the integrated carbon-normalized absorbance from 250 – 450 nm (Helms et al., 2008). The percent loss of CDOM is shown under the label of each leachate.
Section 2. Presence-absence analysis of FT-ICR-MS formulas

Figure S2. Van Krevelen diagrams of bio-labile formulas identified in the four pyDOM samples using a presence-absence approach (Sleighter et al., 2012). The number of formulas and the corresponding percentage (relative to the total number of formulas in the two samples being compared) are shown in the legends. The carbon losses quantified by Bostick et al. (2021) are listed under the legends. The black lines indicate modified aromaticity index cutoffs (AIMOD; Koch and Dittmar, 2006, 2016), and the red box indicates the peptide region (valid only for N-containing formulas).
Figure S3. Van Krevelen diagrams of bio-resistant formulas identified in the four pyDOM samples using a presence-absence approach (Sleighter et al., 2012). The number of formulas and the corresponding percentage (relative to the total number of formulas in the two samples being compared) are shown in the legends. The carbon losses quantified by Bostick et al. (2021) are listed under the legends. The black lines indicate modified aromaticity index cutoffs (Al_{MOD}; Koch and Dittmar, 2006, 2016), and the red box indicates the peptide region (valid only for N-containing formulas).
Figure S4. Van Krevelen diagrams of bio-produced formulas identified in pyDOM samples using a presence-absence approach (Sleighter et al., 2012). The number of formulas and the corresponding percentage (relative to the total number of formulas in the two samples being compared) are shown in the legends. The carbon losses quantified by Bostick et al. (2021) are listed under the legends. The black lines indicate modified aromaticity index cutoffs (AIMOD; Koch and Dittmar, 2006, 2016), and the red box indicates the peptide region (valid only for N-containing formulas).
Section 3. H/C versus molecular weight plots

Figure S5. Hydrogen-to-carbon (H/C) ratio versus molecular weight plots of 10-day microbially incubated pyDOM leachates. Formulas are classified as bio-labile (formulas only found in the control pyDOM leachates) and bio-produced (formulas only found in the bio-incubated samples). Formulas that are present in both the control and bio-incubated samples are operationally classified as bio-resistant and not shown for clarity. These three classes of molecules are separately plotted in Figs. S6-8. The number of formulas in each of these pools is shown in the legends along with their corresponding percentages (relative to total number of formulas in the two samples being compared). The carbon losses quantified by Bostick et al. (2021) are listed under the legends. The red lines indicate where peptide-like formulas would plot (valid only for N-containing formulas).
Figure S6. Hydrogen-to-carbon (H/C) ratio versus molecular weight plots of the bio-labile formulas. The number of formulas in each of these pools is shown in the legends along with their corresponding percentages (relative to total number of formulas in the two samples being compared). The carbon losses quantified by Bostick et al. (2021) are listed under the legends. The red lines indicate where peptide-like formulas would plot (valid only for N-containing formulas).
Figure S7. Hydrogen-to-carbon (H/C) ratio versus molecular weight plots of the bio-resistant formulas. The number of formulas in each of these pools is shown in the legends along with their corresponding percentages (relative to total number of formulas in the two samples being compared). The carbon losses quantified by Bostick et al. (2021) are listed under the legends. The red lines indicate where peptide-like formulas would plot (valid only for N-containing formulas).
Figure S8. Hydrogen-to-carbon (H/C) ratio versus molecular weight plots of the bio-produced formulas. The number of formulas in each of these pools is shown in the legends along with their corresponding percentages (relative to total number of formulas in the two samples being compared). The carbon losses quantified by Bostick et al. (2021) are listed under the legends. The red lines indicate where peptide-like formulas would plot (valid only for N-containing formulas).
Section 4. Evaluation of bio-resistant formulas

Figure S9. Abundance scatterplots of the bio-resistant formulas following Sleighter et al. (2012). This approach evaluates the similarity in relative abundance of each common formula among the control and its corresponding bio-incubated sample. A high $R^2$ value indicates a high similarity in mass spectrometric abundance of these formulas.
Section 5. Comparison of bio-produced formulas with estuarine and marine samples

Bio-produced formulas from the pyDOM incubations were combined into one master formula list (total of 4762 formulas). These formulas were searched in previously published data of aquatic samples to test if bio-incubations of pyDOM produce molecular formulas that have been found across an estuarine transect or in the ocean. Bio-produced formulas that were found common to environmental dissolved organic matter (DOM) were attributed to carboxyl-rich alicyclic molecules (CRAM) if they met the following criteria: DBE/C = 0.30 − 0.68; DBE/H = 0.20 − 0.95; DBE/O = 0.77 − 1.75 (Hertkorn et al., 2006).

Table S1. Overlap of bio-produced formulas of pyDOM with estuarine and marine DOM samples. Sample codes are listed in parentheses in addition to the sample preparation approach (solid-phase extraction using C18 or PPL cartridges; or RO/ED = reverse osmosis/electrodialysis). For the first five samples (estuarine transect of the Elizabeth River, VA, USA), salinity values are listed in square brackets. Number of common formulas are reported in relative to the number of formulas of the environmental sample. Number of CRAM formulas are reported relative to the total number of common formulas.

<table>
<thead>
<tr>
<th>Sample Name</th>
<th>Number of Formulas</th>
<th>Number of estuarine/marine formulas that can be produced by bio-incubation of pyDOM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dismal Swamp [0], C18³</td>
<td>1752</td>
<td>223 (13 %) CRAM: 123 (55%)</td>
</tr>
<tr>
<td>Great Bridge [11], C18³</td>
<td>1727</td>
<td>292 (17 %) CRAM: 148 (51%)</td>
</tr>
<tr>
<td>Town Point [20], C18³</td>
<td>1303</td>
<td>228 (18 %) CRAM: 102 (45%)</td>
</tr>
<tr>
<td>Chesapeake Bay Bridge [22], C18³</td>
<td>1079</td>
<td>193 (18 %) CRAM: 81 (42%)</td>
</tr>
<tr>
<td>Off Shore Coast [32], C18³</td>
<td>1189</td>
<td>212 (18 %) CRAM: 88 (42%)</td>
</tr>
<tr>
<td>N. Atlantic Ocean surface water (DOM411), PPL³</td>
<td>2402</td>
<td>227 (10 %) CRAM: 159 (70%)</td>
</tr>
<tr>
<td>N. Atlantic Ocean surface water (DOM412), PPL³</td>
<td>3524</td>
<td>289 (8 %) CRAM: 192 (66%)</td>
</tr>
<tr>
<td>N. Atlantic Ocean surface water (DOM417), PPL³</td>
<td>3312</td>
<td>263 (8 %) CRAM: 181 (69%)</td>
</tr>
<tr>
<td>N. Pacific Ocean surface water (DOM 1), RO/ED³</td>
<td>1918</td>
<td>261 (14 %) CRAM: 155 (59%)</td>
</tr>
<tr>
<td>N. Pacific Ocean surface water (DOM 1r), RO/ED³</td>
<td>1950</td>
<td>258 (13 %) CRAM: 152 (59%)</td>
</tr>
<tr>
<td>N. Atlantic Ocean abyssal water (DOM 2), RO/ED³</td>
<td>1697</td>
<td>284 (17 %) CRAM: 154 (54%)</td>
</tr>
<tr>
<td>N. Atlantic Ocean abyssal water (DOM 2r), RO/ED³</td>
<td>1756</td>
<td>308 (18 %) CRAM: 167 (54%)</td>
</tr>
<tr>
<td>Coastal N. Pacific Ocean water (DOM 3), PPL³</td>
<td>2226</td>
<td>265 (12 %) CRAM: 161 (61%)</td>
</tr>
<tr>
<td>Coastal N. Pacific Ocean water (DOM 3 rep), PPL³</td>
<td>2256</td>
<td>278 (12 %) CRAM: 168 (60%)</td>
</tr>
<tr>
<td>Coastal N. Pacific Ocean water (DOM 4), PPL³</td>
<td>2325</td>
<td>287 (12 %) CRAM: 169 (59%)</td>
</tr>
<tr>
<td>Coastal N. Pacific Ocean water (DOM 4 rep), PPL³</td>
<td>2429</td>
<td>288 (12 %) CRAM: 177 (61%)</td>
</tr>
</tbody>
</table>

³Estuarine transect data from Sleighter and Hatcher (2008). Only the Off Shore Coast sample is considered marine.

Unpublished data from samples obtained during the WACS-2 cruise (R/V Knorr) as part of the Western Atlantic Climate Study (WACS).

Chen et al. (2014)

Sleighter et al. (2012)

The four different pyDOM samples were also individually compared to all marine DOM samples combined together in a master environmental formula list (only the Off Shore Coast sample from the estuarine transect was used). Oak 400 Fresh had 265 formulas found in oceanic DOM, Oak 400 Photo: 157 formulas, Oak 650 Fresh: 121 formulas, Oak 650 Photo: 173 formulas. This indicated a variable potential of pyDOM to be microbially transformed into marine-like DOM.
Section 6. Analysis of variance (ANOVA) of bio-produced peptide-like formulas

Table S2. Molecular metrics of peptide-like bio-produced formulas (N-containing, 1.5 ≤ H/C ≤ 2.0, 0.1 ≤ O/C ≤ 0.67) found in pyDOM and sucrose samples after the 10-day incubation. The metrics below are reported as number-weighed mean ± standard deviation. The molecular metrics colored in red correspond to the means that were found to be significantly different (p < 0.05, ANOVA followed by Scheffé’s post-hoc test) from at least one of the other four means.

<table>
<thead>
<tr>
<th></th>
<th>Oak 400 Fresh</th>
<th>Oak 400 Photo</th>
<th>Oak 650 Fresh</th>
<th>Oak 650 Photo</th>
<th>Sucrose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of bio-produced formulas</td>
<td>1778</td>
<td>1111</td>
<td>1246</td>
<td>1456</td>
<td>1339</td>
</tr>
<tr>
<td>Number of peptide-like bio-produced formulas</td>
<td>541 (30%)</td>
<td>261 (23%)</td>
<td>497 (40%)</td>
<td>314 (22%)</td>
<td>160 (12%)</td>
</tr>
<tr>
<td>Number of identified oligopeptides</td>
<td>14</td>
<td>5</td>
<td>11</td>
<td>18</td>
<td>2</td>
</tr>
<tr>
<td>C number</td>
<td>28.5 ± 7.6</td>
<td>30.9 ± 10.9</td>
<td>30.7 ± 7.6</td>
<td>30.3 ± 8.7</td>
<td>31.7 ± 9.6</td>
</tr>
<tr>
<td>H number</td>
<td>49.8 ± 14.4</td>
<td>54 ± 20.6</td>
<td>53.7 ± 14.8</td>
<td>54 ± 16.5</td>
<td>55.4 ± 18.5</td>
</tr>
<tr>
<td>O number</td>
<td>7.8 ± 2.6</td>
<td>7.8 ± 3.2</td>
<td>7.8 ± 2.9</td>
<td>9.0 ± 2.8</td>
<td>7.9 ± 3.1</td>
</tr>
<tr>
<td>N number</td>
<td>2.4 ± 1.1</td>
<td>2.8 ± 1.3</td>
<td>2.5 ± 1.2</td>
<td>2.4 ± 1.2</td>
<td>2.4 ± 1.3</td>
</tr>
<tr>
<td>O/C ratio</td>
<td>0.28 ± 0.08</td>
<td>0.26 ± 0.09</td>
<td>0.25 ± 0.08</td>
<td>0.31 ± 0.10</td>
<td>0.25 ± 0.08</td>
</tr>
<tr>
<td>H/C ratio</td>
<td>1.74 ± 0.12</td>
<td>1.74 ± 0.13</td>
<td>1.74 ± 0.13</td>
<td>1.78 ± 0.16</td>
<td>1.74 ± 0.14</td>
</tr>
<tr>
<td>N/C ratio</td>
<td>0.085 ± 0.037</td>
<td>0.094 ± 0.045</td>
<td>0.082 ± 0.038</td>
<td>0.083 ± 0.045</td>
<td>0.078 ± 0.042</td>
</tr>
<tr>
<td>H/N ratio</td>
<td>24.8 ± 11.4</td>
<td>23.5 ± 13.4</td>
<td>26 ± 13.2</td>
<td>28.6 ± 16.7</td>
<td>29.4 ± 16</td>
</tr>
<tr>
<td>O/N ratio</td>
<td>4.0 ± 2.2</td>
<td>3.5 ± 2.2</td>
<td>3.8 ± 2.5</td>
<td>5.1 ± 3.5</td>
<td>4.3 ± 2.7</td>
</tr>
<tr>
<td>MW&lt;sup&gt;a&lt;/sup&gt;</td>
<td>550 ± 140</td>
<td>589 ± 188</td>
<td>582 ± 147</td>
<td>596 ± 143</td>
<td>597 ± 172</td>
</tr>
<tr>
<td>DBE&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5.81 ± 1.78</td>
<td>6.28 ± 2.17</td>
<td>6.13 ± 2.06</td>
<td>5.51 ± 2.59</td>
<td>6.2 ± 2.33</td>
</tr>
<tr>
<td>DBE/C&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.211 ± 0.065</td>
<td>0.215 ± 0.071</td>
<td>0.206 ± 0.069</td>
<td>0.189 ± 0.083</td>
<td>0.203 ± 0.071</td>
</tr>
<tr>
<td>DBE-O&lt;sup&gt;d&lt;/sup&gt;</td>
<td>-2.27 ± 2.75</td>
<td>-1.75 ± 3.52</td>
<td>-1.90 ± 3.55</td>
<td>-3.82 ± 4.26</td>
<td>-1.86 ± 3.65</td>
</tr>
<tr>
<td>AI&lt;sub&gt;MOD&lt;/sub&gt;&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.077 ± 0.05</td>
<td>0.090 ± 0.052</td>
<td>0.083 ± 0.049</td>
<td>0.089 ± 0.057</td>
<td>0.116 ± 0.049</td>
</tr>
<tr>
<td>NOSC&lt;sup&gt;f&lt;/sup&gt;</td>
<td>-0.929 ± 0.239</td>
<td>-0.933 ± 0.259</td>
<td>-0.984 ± 0.227</td>
<td>-0.903 ± 0.269</td>
<td>-1.002 ± 0.218</td>
</tr>
</tbody>
</table>

<sup>a</sup>Molecular Weight (Da), <sup>b</sup>Double-bond equivalency, <sup>c</sup>Carbon-normalized DBE, <sup>d</sup>Oxygen-corrected DBE, <sup>e</sup>Modified Aromaticity Index, <sup>f</sup>Nominal Oxidation State of Carbon

The proteinaceous formulas in the four samples were evaluated using one-way ANOVA to assess the variability in their composition. Peptide-like formulas seemed similar when plotted in the van Krevelen space (Figs. 1 and S5). To further assess them, different molecular parameters were derived from their formula lists – average number of elements (C, H, O, N), elemental ratios (O/C, H/C, N/C, H/N, O/N), molecular weight, double-bond equivalencies (DBE, DBE/C, DBE-O), modified aromaticity index (AI<sub>MOD</sub>) and nominal oxidation state of carbon (NOSC). When each metric was evaluated using ANOVA, there was at least one sample among the five being compared that had a significantly different mean. Using Scheffé's post-hoc test, it was observed that it was not the same sample that was statistically different each time, which indicated that the bio-produced peptide-like molecules were of vast diversity among the different incubations.
## Section 7. Oligopeptide Sequences

### Table S3. Oligopeptide sequences consistent with bio-produced formulas of each pyDOM sample. *Combinations can be of any order*

<table>
<thead>
<tr>
<th>Sample</th>
<th>Measured m/z</th>
<th>Amino Acid Combination#</th>
<th>Molecular Weight</th>
<th>Molecular Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oak 400 Fresh</td>
<td>201.1246</td>
<td>AL</td>
<td>202.1317</td>
<td>C₇H₁₈O₃N₂</td>
</tr>
<tr>
<td>Oak 400 Fresh</td>
<td>356.2192</td>
<td>OLL</td>
<td>357.2264</td>
<td>C₁₇H₃₁O₇N₃</td>
</tr>
<tr>
<td>Oak 400 Fresh</td>
<td>455.2874</td>
<td>OLLV</td>
<td>456.2948</td>
<td>C₂₃H₄₀O₆N₄</td>
</tr>
<tr>
<td>Oak 400 Fresh</td>
<td>512.3457</td>
<td>ALLVVV</td>
<td>513.3526</td>
<td>C₂₅H₄₇O₆N₅</td>
</tr>
<tr>
<td>Oak 400 Fresh</td>
<td>512.3457</td>
<td>GLLLL</td>
<td>513.3526</td>
<td>C₂₅H₴₇O₆N₅</td>
</tr>
<tr>
<td>Oak 400 Fresh</td>
<td>512.3457</td>
<td>VVVVV</td>
<td>513.3526</td>
<td>C₂₅H₴₇O₆N₅</td>
</tr>
<tr>
<td>Oak 400 Fresh</td>
<td>514.3251</td>
<td>ALLLS</td>
<td>515.3319</td>
<td>C₂₄H₴₅O₇N₅</td>
</tr>
<tr>
<td>Oak 400 Fresh</td>
<td>514.3251</td>
<td>ALTTV</td>
<td>515.3319</td>
<td>C₂₄H₴₅O₇N₅</td>
</tr>
<tr>
<td>Oak 400 Fresh</td>
<td>514.3251</td>
<td>GLLLT</td>
<td>515.3319</td>
<td>C₂₄H₴₅O₇N₅</td>
</tr>
<tr>
<td>Oak 400 Fresh</td>
<td>514.3251</td>
<td>LSVVVV</td>
<td>515.3319</td>
<td>C₂₄H₴₅O₇N₅</td>
</tr>
<tr>
<td>Oak 400 Fresh</td>
<td>514.3251</td>
<td>TVVVVV</td>
<td>515.3319</td>
<td>C₂₄H₴₅O₇N₅</td>
</tr>
<tr>
<td>Oak 400 Fresh</td>
<td>526.3607</td>
<td>ALLLV</td>
<td>527.3683</td>
<td>C₂₆H₴₉O₆N₅</td>
</tr>
<tr>
<td>Oak 400 Fresh</td>
<td>526.3607</td>
<td>GLLLL</td>
<td>527.3683</td>
<td>C₂₆H₴₉O₆N₅</td>
</tr>
<tr>
<td>Oak 400 Fresh</td>
<td>526.3607</td>
<td>LVVVV</td>
<td>527.3683</td>
<td>C₂₆H₴₉O₆N₅</td>
</tr>
<tr>
<td>Oak 400 Photo</td>
<td>341.2195</td>
<td>LPX</td>
<td>342.2267</td>
<td>C₁₆H₃₀O₄N₄</td>
</tr>
<tr>
<td>Oak 400 Photo</td>
<td>341.2195</td>
<td>KPV</td>
<td>342.2267</td>
<td>C₁₆H₃₀O₄N₄</td>
</tr>
<tr>
<td>Oak 400 Photo</td>
<td>350.1836</td>
<td>HPV</td>
<td>351.1907</td>
<td>C₁₆H₂₅O₄N₅</td>
</tr>
<tr>
<td>Oak 400 Photo</td>
<td>528.3188</td>
<td>LLVVW</td>
<td>529.3264</td>
<td>C₂₈H₴₃O₅N₅</td>
</tr>
<tr>
<td>Oak 400 Photo</td>
<td>552.3768</td>
<td>LLLLPV</td>
<td>553.3839</td>
<td>C₂₈H₵₁O₆N₅</td>
</tr>
<tr>
<td>Oak 650 Fresh</td>
<td>498.3293</td>
<td>AALLLL</td>
<td>499.3370</td>
<td>C₂₄H₴₅O₆N₅</td>
</tr>
<tr>
<td>Oak 650 Fresh</td>
<td>498.3293</td>
<td>ALVVVV</td>
<td>499.3370</td>
<td>C₂₄H₴₅O₆N₅</td>
</tr>
<tr>
<td>Oak 650 Fresh</td>
<td>498.3293</td>
<td>GLVVV</td>
<td>499.3370</td>
<td>C₂₄H₴₅O₆N₅</td>
</tr>
<tr>
<td>Oak 650 Fresh</td>
<td>512.3455</td>
<td>ALLVV</td>
<td>513.3526</td>
<td>C₂₅H₴₇O₆N₅</td>
</tr>
<tr>
<td>Oak 650 Fresh</td>
<td>512.3455</td>
<td>GLLLL</td>
<td>513.3526</td>
<td>C₂₅H₴₇O₆N₅</td>
</tr>
<tr>
<td>Oak 650 Fresh</td>
<td>512.3455</td>
<td>VVVVV</td>
<td>513.3526</td>
<td>C₂₅H₴₇O₆N₅</td>
</tr>
<tr>
<td>Oak 650 Fresh</td>
<td>512.3455</td>
<td>VVVVV</td>
<td>513.3526</td>
<td>C₂₅H₴₇O₆N₅</td>
</tr>
<tr>
<td>Oak 650 Fresh</td>
<td>552.3042</td>
<td>DLLPP</td>
<td>553.3112</td>
<td>C₂₆H₴₃O₈N₅</td>
</tr>
<tr>
<td>Oak 650 Fresh</td>
<td>552.3042</td>
<td>ELPPV</td>
<td>553.3112</td>
<td>C₂₆H₴₃O₈N₅</td>
</tr>
<tr>
<td>Oak 650 Fresh</td>
<td>552.3042</td>
<td>OOLPV</td>
<td>553.3112</td>
<td>C₂₆H₴₃O₈N₅</td>
</tr>
<tr>
<td>Oak 650 Fresh</td>
<td>552.3042</td>
<td>OLUVV</td>
<td>553.3112</td>
<td>C₂₆H₴₃O₈N₅</td>
</tr>
<tr>
<td>Oak 650 Fresh</td>
<td>552.3042</td>
<td>LLPUT</td>
<td>553.3112</td>
<td>C₂₆H₴₃O₈N₅</td>
</tr>
<tr>
<td>Oak 650 Photo</td>
<td>242.1508</td>
<td>KP</td>
<td>243.1583</td>
<td>C₁₁H₂₁O₃N₃</td>
</tr>
<tr>
<td>Oak 650 Photo</td>
<td>342.2034</td>
<td>OLV</td>
<td>343.2107</td>
<td>C₁₆H₂₉O₅N₃</td>
</tr>
<tr>
<td>Oak 650 Photo</td>
<td>356.2190</td>
<td>OLL</td>
<td>357.2264</td>
<td>C₁₇H₳₁O₅N₃</td>
</tr>
<tr>
<td>Oak 650 Photo</td>
<td>552.2676</td>
<td>ALSTY</td>
<td>553.2748</td>
<td>C₂₅H₹₃O₉N₅</td>
</tr>
<tr>
<td>Oak 650 Photo</td>
<td>552.2676</td>
<td>ATTYV</td>
<td>553.2748</td>
<td>C₂₅H₹₃O₉N₅</td>
</tr>
<tr>
<td>Compound</td>
<td>M/z</td>
<td>MRM Transition</td>
<td>M/z</td>
<td>Molecular Formula</td>
</tr>
<tr>
<td>------------------</td>
<td>-------</td>
<td>----------------</td>
<td>-------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Oak 650 Photo</td>
<td>552.2676</td>
<td>DOLPP</td>
<td>553.2748</td>
<td>C$<em>{25}$H$</em>{39}$O$_9$N$_5$</td>
</tr>
<tr>
<td>Oak 650 Photo</td>
<td>552.2676</td>
<td>DLPUV</td>
<td>553.2748</td>
<td>C$<em>{25}$H$</em>{39}$O$_9$N$_5$</td>
</tr>
<tr>
<td>Oak 650 Photo</td>
<td>552.2676</td>
<td>EOPPV</td>
<td>553.2748</td>
<td>C$<em>{25}$H$</em>{39}$O$_9$N$_5$</td>
</tr>
<tr>
<td>Oak 650 Photo</td>
<td>552.2676</td>
<td>EPUVV</td>
<td>553.2748</td>
<td>C$<em>{25}$H$</em>{39}$O$_9$N$_5$</td>
</tr>
<tr>
<td>Oak 650 Photo</td>
<td>552.2676</td>
<td>GLTTY</td>
<td>553.2748</td>
<td>C$<em>{25}$H$</em>{39}$O$_9$N$_5$</td>
</tr>
<tr>
<td>Oak 650 Photo</td>
<td>552.2676</td>
<td>OOPPV</td>
<td>553.2748</td>
<td>C$<em>{25}$H$</em>{39}$O$_9$N$_5$</td>
</tr>
<tr>
<td>Oak 650 Photo</td>
<td>552.2676</td>
<td>OOUVV</td>
<td>553.2748</td>
<td>C$<em>{25}$H$</em>{39}$O$_9$N$_5$</td>
</tr>
<tr>
<td>Oak 650 Photo</td>
<td>552.2676</td>
<td>OLPUT</td>
<td>553.2748</td>
<td>C$<em>{25}$H$</em>{39}$O$_9$N$_5$</td>
</tr>
<tr>
<td>Oak 650 Photo</td>
<td>552.2676</td>
<td>LLUUS</td>
<td>553.2748</td>
<td>C$<em>{25}$H$</em>{39}$O$_9$N$_5$</td>
</tr>
<tr>
<td>Oak 650 Photo</td>
<td>552.2676</td>
<td>LFSST</td>
<td>553.2748</td>
<td>C$<em>{25}$H$</em>{39}$O$_9$N$_5$</td>
</tr>
<tr>
<td>Oak 650 Photo</td>
<td>552.2676</td>
<td>LUUTV</td>
<td>553.2748</td>
<td>C$<em>{25}$H$</em>{39}$O$_9$N$_5$</td>
</tr>
<tr>
<td>Oak 650 Photo</td>
<td>552.2676</td>
<td>FSTTV</td>
<td>553.2748</td>
<td>C$<em>{25}$H$</em>{39}$O$_9$N$_5$</td>
</tr>
<tr>
<td>Oak 650 Photo</td>
<td>552.2676</td>
<td>SSYVV</td>
<td>553.2748</td>
<td>C$<em>{25}$H$</em>{39}$O$_9$N$_5$</td>
</tr>
<tr>
<td>Sucrose</td>
<td>340.1880</td>
<td>OLP</td>
<td>341.1951</td>
<td>C$<em>{16}$H$</em>{27}$O$_5$N$_3$</td>
</tr>
<tr>
<td>Sucrose</td>
<td>340.1880</td>
<td>LUV</td>
<td>341.1951</td>
<td>C$<em>{16}$H$</em>{27}$O$_5$N$_3$</td>
</tr>
</tbody>
</table>
**Section 8.** Kendrick mass defect plots of bio-produced formulas

**Figure S10.** Kendrick mass defect (KMD) versus Kendrick nominal mass plots for the oxygen (O) series within the bio-produced formulas of the four pyDOM samples. Formulas not part of the O KMD series are colored in gray. Formulas in **dark green** are substrates, with their oxygenation products colored in **light green**. The number of formulas of each of these pools are shown in the legends along with corresponding percentages (relative to total number of bio-produced formulas).
Figure S11. Kendrick mass defect (KMD) versus Kendrick nominal mass plots for the carbonyl (CO) series within the bio-produced formulas of the four pyDOM samples. Formulas not part of the CO KMD series are colored in gray. Formulas in dark green are substrates, with their oxygenation products colored in light green. The number of formulas of each of these pools are shown in the legends along with corresponding percentages (relative to total number of bio-produced formulas).
Figure S12. Kendrick mass defect (KMD) versus Kendrick nominal mass plots for the carboxyl (COO) series within the bio-produced formulas of the four pyDOM samples. Formulas not part of the COO KMD series are colored in gray. Formulas in dark green are substrates, with their oxygenation products colored in light green. The number of formulas of each of these pools are shown in the legends along with corresponding percentages (relative to total number of bio-produced formulas).
Section 9. Correlation analysis of molecular diversity and NMR data

Table S4. Data used for Pearson correlations between molecular diversity (number of FT-ICR-MS molecular formulas) and functional group content from 1D NMR (Bostick et al., 2021). $R^2$ and p-values are listed for each functional group in the corresponding color.

<table>
<thead>
<tr>
<th>Functional group</th>
<th>Oak 400 Fresh</th>
<th>Oak 400 Photo</th>
<th>Oak 650 Fresh</th>
<th>Oak 650 Photo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of bio-labile formulas</td>
<td>1646</td>
<td>1242</td>
<td>1364</td>
<td>1410</td>
</tr>
<tr>
<td>Number of bio-produced formulas</td>
<td>1778</td>
<td>1111</td>
<td>1246</td>
<td>1456</td>
</tr>
<tr>
<td>Aldehyde (O=CH)</td>
<td>1778</td>
<td>1111</td>
<td>1246</td>
<td>1456</td>
</tr>
<tr>
<td>$R^2 = 0.1263$, $R^2 = 0.2374$</td>
<td>3.18%</td>
<td>4.52%</td>
<td>10.99%</td>
<td>4.24%</td>
</tr>
<tr>
<td>$p = 0.6448$, $p = 0.5130$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aryl</td>
<td>1778</td>
<td>1111</td>
<td>1246</td>
<td>1456</td>
</tr>
<tr>
<td>$R^2 = 0.0094$, $R^2 = 0.0668$</td>
<td>9.87%</td>
<td>8.47%</td>
<td>20.65%</td>
<td>7.54%</td>
</tr>
<tr>
<td>$p = 0.9031$, $p = 0.7418$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Olefinic (C=C)</td>
<td>1778</td>
<td>1111</td>
<td>1246</td>
<td>1456</td>
</tr>
<tr>
<td>$R^2 = 0.9472$, $R^2 = 0.9978$</td>
<td>7.64%</td>
<td>15.60%</td>
<td>14.31%</td>
<td>11.41%</td>
</tr>
<tr>
<td>$p = 0.0267$, $p = 0.0011$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HC-O-R</td>
<td>1778</td>
<td>1111</td>
<td>1246</td>
<td>1456</td>
</tr>
<tr>
<td>$R^2 = 0.4217$, $R^2 = 0.3385$</td>
<td>6.75%</td>
<td>23.64%</td>
<td>4.57%</td>
<td>9.41%</td>
</tr>
<tr>
<td>$p = 0.3509$, $p = 0.4184$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HC-C=Y</td>
<td>1778</td>
<td>1111</td>
<td>1246</td>
<td>1456</td>
</tr>
<tr>
<td>$R^2 = 0.0201$, $R^2 = 0.0511$</td>
<td>12.33%</td>
<td>13.14%</td>
<td>4.49%</td>
<td>9.13%</td>
</tr>
<tr>
<td>$p = 0.8590$, $p = 0.7746$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HC-C-C-X</td>
<td>1778</td>
<td>1111</td>
<td>1246</td>
<td>1456</td>
</tr>
<tr>
<td>$R^2 = 0.4639$, $R^2 = 0.3968$</td>
<td>3.98%</td>
<td>5.99%</td>
<td>6.52%</td>
<td>7.38%</td>
</tr>
<tr>
<td>$p = 0.3201$, $p = 0.3930$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methylene (CH$_2$)</td>
<td>1778</td>
<td>1111</td>
<td>1246</td>
<td>1456</td>
</tr>
<tr>
<td>$R^2 = 0.1287$, $R^2 = 0.0997$</td>
<td>6.46%</td>
<td>7.85%</td>
<td>11.57%</td>
<td>12.65%</td>
</tr>
<tr>
<td>$p = 0.6405$, $p = 0.6836$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methyl (CH$_3$)</td>
<td>1778</td>
<td>1111</td>
<td>1246</td>
<td>1456</td>
</tr>
<tr>
<td>$R^2 = 0.0653$, $R^2 = 0.1664$</td>
<td>0.89%</td>
<td>0.84%</td>
<td>0.25%</td>
<td>0.93%</td>
</tr>
<tr>
<td>$p = 0.7454$, $p = 0.5926$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formate (HCOO$^-$)</td>
<td>1778</td>
<td>1111</td>
<td>1246</td>
<td>1456</td>
</tr>
<tr>
<td>$R^2 = 0.0033$, $R^2 = 0.0124$</td>
<td>10.57%</td>
<td>3.51%</td>
<td>24.18%</td>
<td>33.91%</td>
</tr>
<tr>
<td>$p = 0.9428$, $p = 0.8889$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methanol (CH$_3$OH)</td>
<td>1778</td>
<td>1111</td>
<td>1246</td>
<td>1456</td>
</tr>
<tr>
<td>$R^2 = 0.9418$, $R^2 = 0.9279$</td>
<td>3.69%</td>
<td>0.47%</td>
<td>0.72%</td>
<td>1.31%</td>
</tr>
<tr>
<td>$p = 0.0297$, $p = 0.0365$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acetate (CH$_3$COO$^-$)</td>
<td>1778</td>
<td>1111</td>
<td>1246</td>
<td>1456</td>
</tr>
<tr>
<td>$R^2 = 0.4217$, $R^2 = 0.3909$</td>
<td>34.63%</td>
<td>15.97%</td>
<td>1.75%</td>
<td>2.10%</td>
</tr>
<tr>
<td>$p = 0.3506$, $p = 0.3748$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


