Labilization and diversification of pyrogenic dissolved organic matter by microbes

Aleksandar I. Goranov[1,*], Andrew S. Wozniak[2], Kyle W. Bostick[3,†], Andrew R. Zimmerman[3], Siddhartha Mitra[4], and Patrick G. Hatcher[1]

[1]Department of Chemistry and Biochemistry, Old Dominion University, Norfolk, VA, USA
[2]School of Marine Science and Policy, College of Earth, Ocean, and Environment, University of Delaware, Lewes, DE, USA
[3]Department of Geological Sciences, University of Florida, Gainesville, FL, USA
[4]Department of Geological Sciences, East Carolina University, Greenville, NC, USA

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

Supplement

Table of Contents

Section 1. Chromophoric dissolved organic matter (CDOM) (Fig. S1) 2

Section 2. Presence~~/ Absence~~ absence analysis of FT-ICR-MS ~~data~~ formulas (Figs. S2-4) 3

Section 3. H/C versus ~~Molecular~~ nmolecular ~~Weight~~ weight ~~analysis~~ plots (Figs. S5-8) 11

Section 4. Evaluation of b~~B~~io-resistant formulas ~~evaluation~~ (Fig. S9) 21

Section 5. Comparison of bio-produced formulas with marine DOM samples (Table S1) 22

Section 6. Analysis of ~~Variance~~ variance (ANOVA) of bio-produced peptide-like ~~organic matter~~formulas (Table S2) 23

Section 7. Peptide sequences (Table S3) 24

Section 8. Kendrick ~~Mass~~ mass ~~Defect~~ defect ~~analysis~~ plots of bio-produced formulas (Fig. S10-12) 26

Section 9. Correlation analysis of molecular diversity and NMR data (Table S4) 30

References 31

[*] Current Address: Department of Earth and Environmental Sciences, Rensselaer Polytechnic Institute, Troy, NY, USA
[†] Current Address: Fugro GeoServices, 6100 Hillcroft Avenue, Houston, TX, USA

Section 1. Chromophoric dissolved organic matter (CDOM)

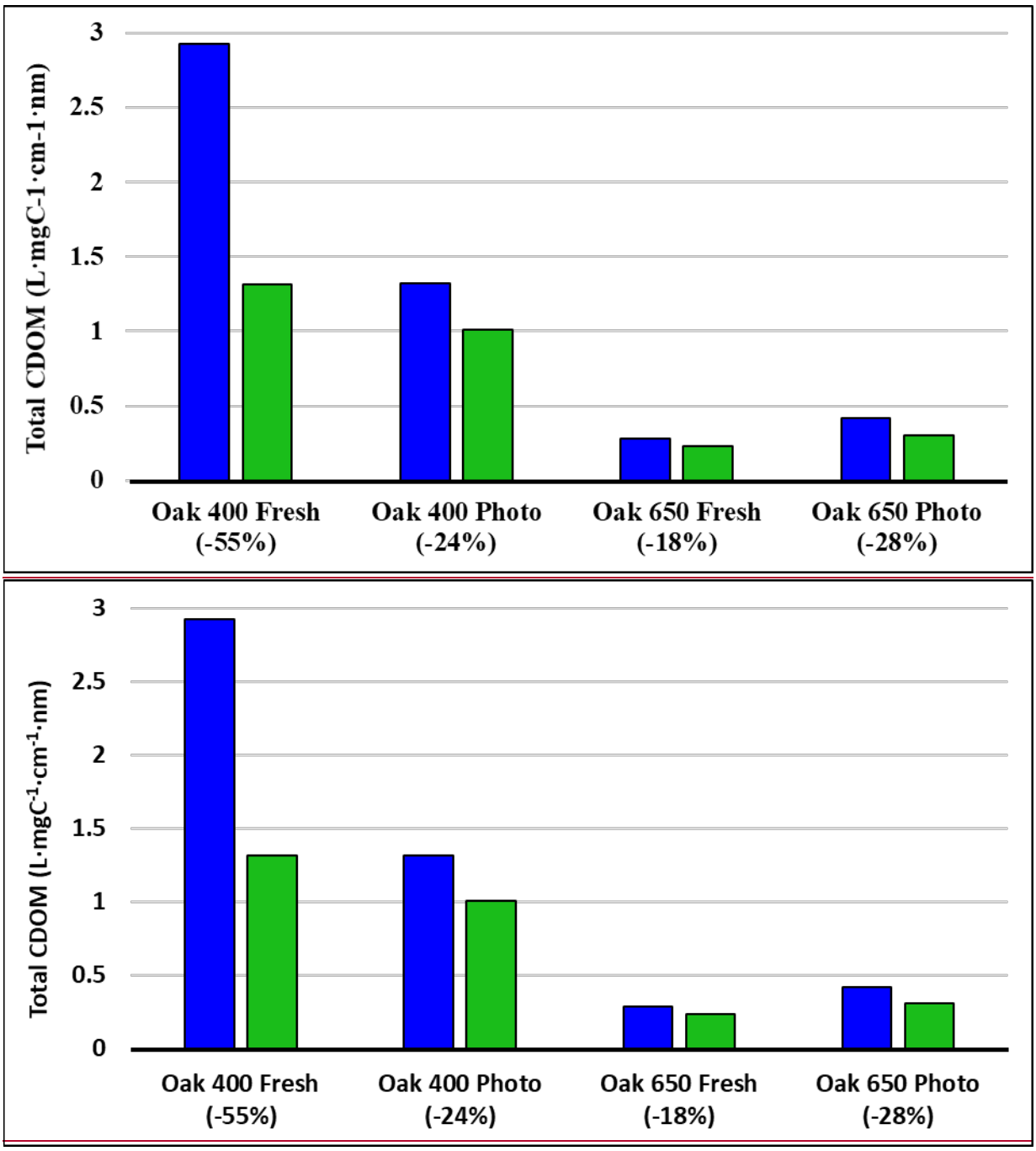
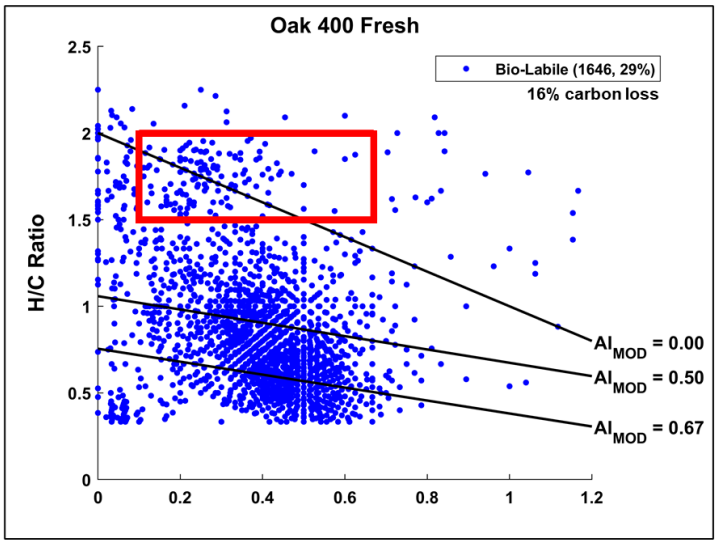
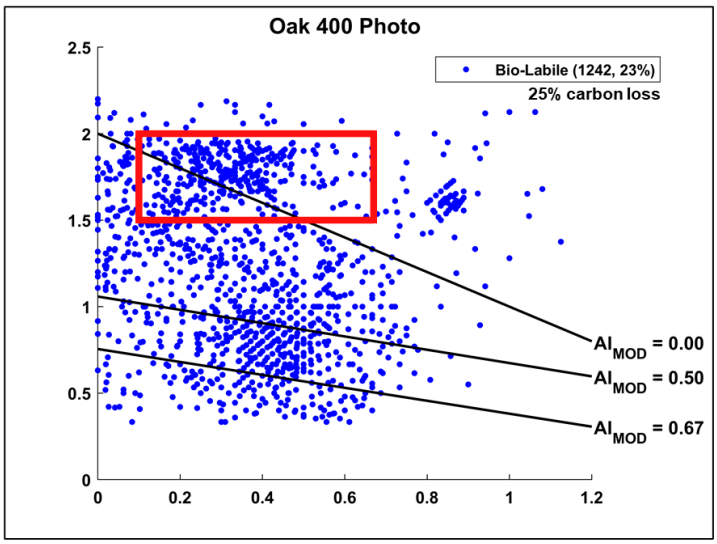
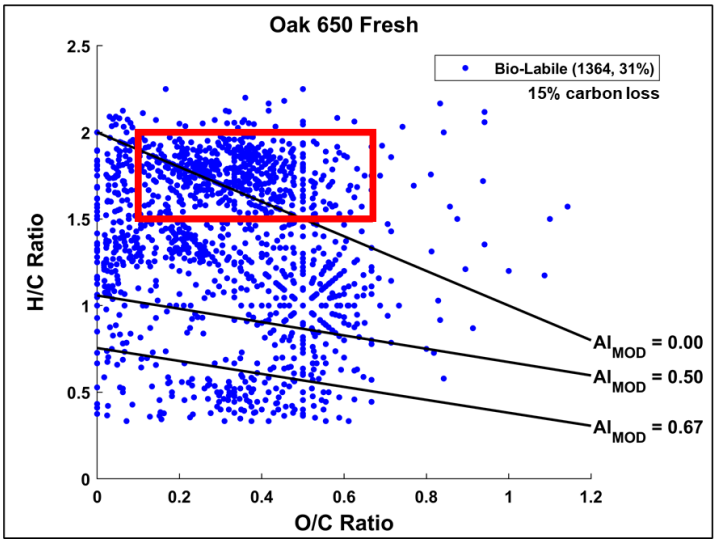


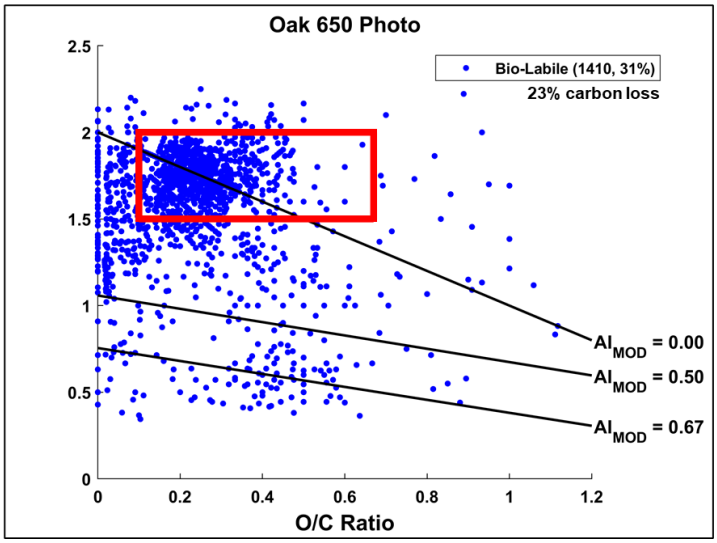
Figure S1. Total chromophoric ~~(colored)~~ dissolved organic matter (CDOM) content of pyDOM leachates before (**blue**) and after (**green**) 10-day biotic incubations. Total CDOM content is reported as the integrated carbon-normalized absorbance from 250 – 450 nm (Helms et al., 2008). The percent loss of CDOM ~~for each leachate~~ is shown ~~as percentage~~ under the label of each leachate.

Section 2. Presence~~/-Absence~~ absence analysis of FT-ICR-MS ~~data~~formulas


Oak 400 Fresh
Bio-Labile (1646, 29%)
16% carbon loss
H/C Ratio
AI_MOD = 0.00
AI_MOD = 0.50
AI_MOD = 0.67


Oak 400 Photo
Bio-Labile (1242, 23%)
25% carbon loss
AI_MOD = 0.00
AI_MOD = 0.50
AI_MOD = 0.67


Oak 650 Fresh
Bio-Labile (1364, 31%)
15% carbon loss
H/C Ratio
O/C Ratio
AI_MOD = 0.00
AI_MOD = 0.50
AI_MOD = 0.67


Oak 650 Photo
Bio-Labile (1410, 31%)
23% carbon loss
O/C Ratio
AI_MOD = 0.00
AI_MOD = 0.50
AI_MOD = 0.67

69

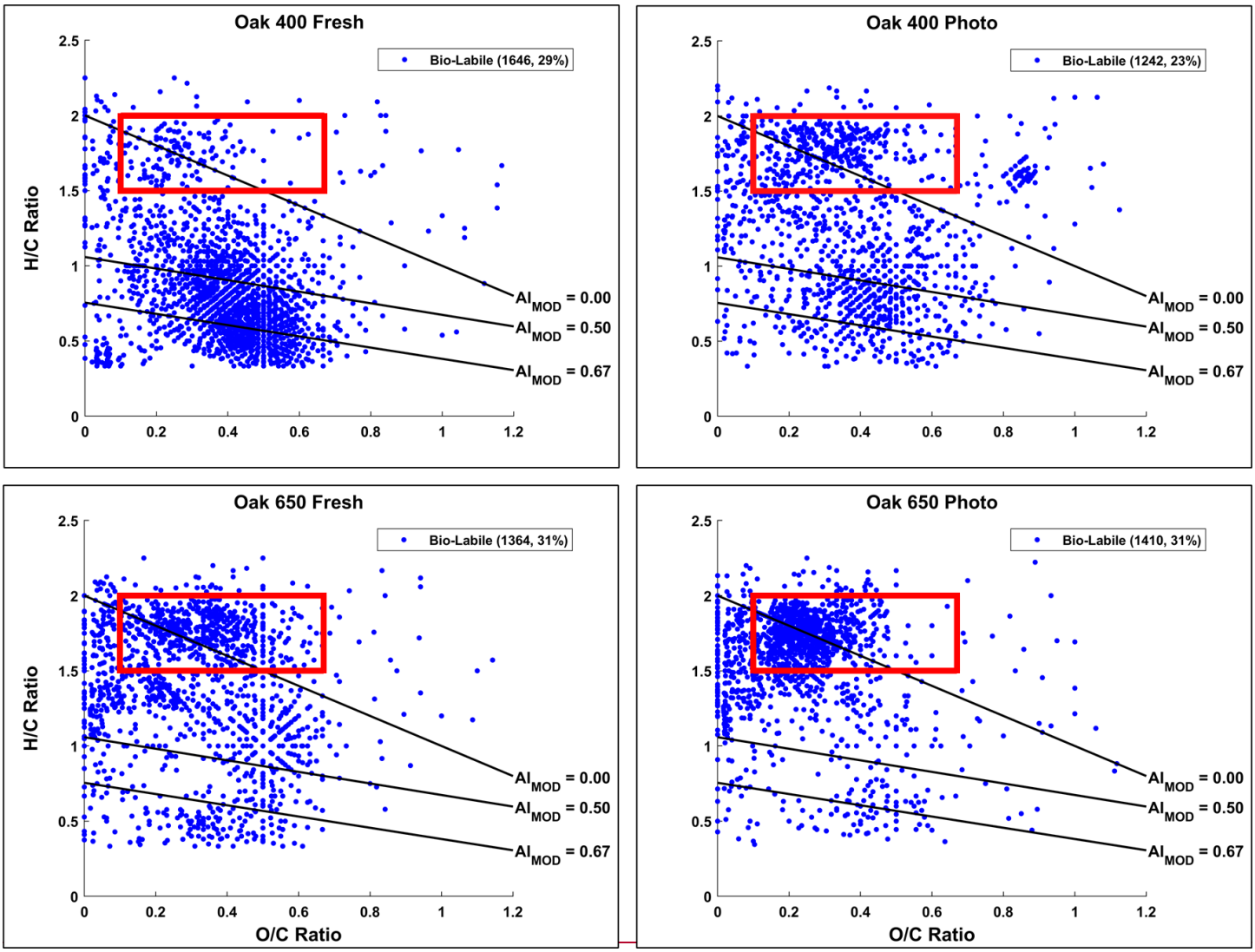
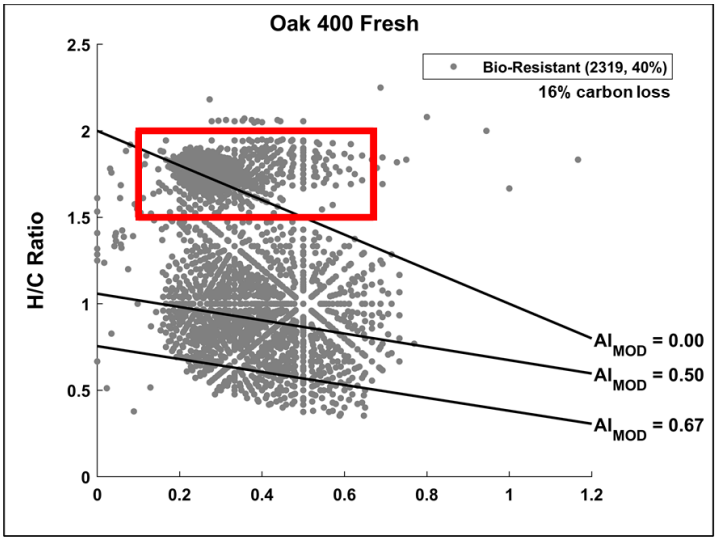
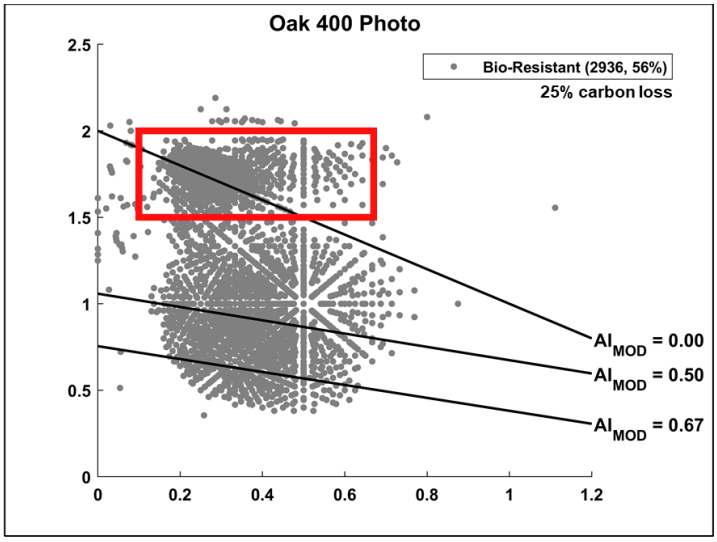
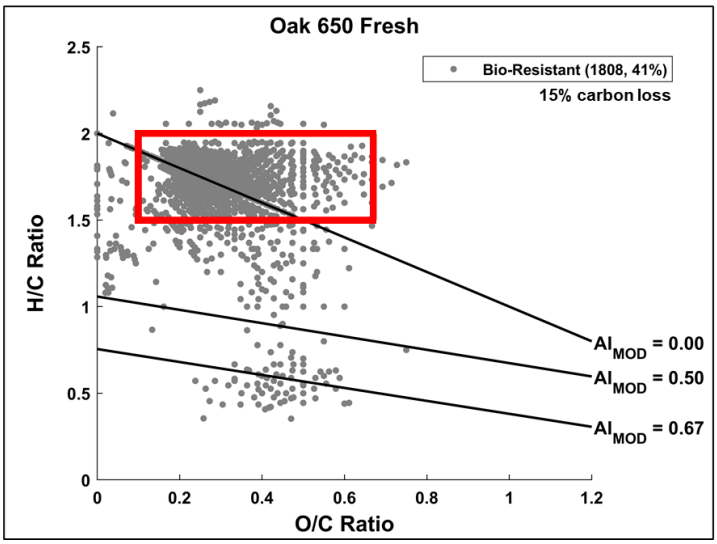
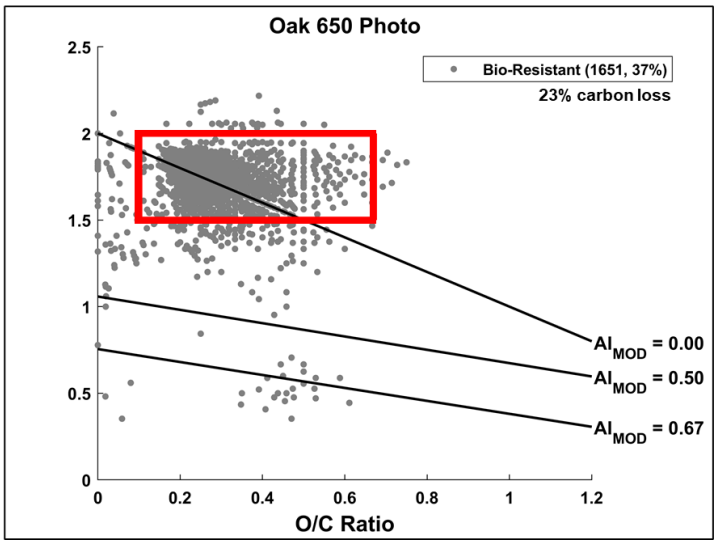


Figure S2. Van Krevelen diagrams of **bio-labile** formulas identified in the four pyDOM samples using a presence~~/~~-absence approach (Sleighter et al., 2012). The number of formulas and the corresponding percentage (relative to the total number of formulas ~~in the two samples being compared~~found in both samples) 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 (AI_{MOD}; Koch and Dittmar, 2006, 2016), and the **red** box indicates the peptide region (valid only for N-containing formulas).


Oak 400 Fresh
Bio-Resistant (2319, 40%)
16% carbon loss
H/C Ratio
AI_{MOD} = 0.00
AI_{MOD} = 0.50
AI_{MOD} = 0.67


Oak 400 Photo
Bio-Resistant (2936, 56%)
25% carbon loss
AI_{MOD} = 0.00
AI_{MOD} = 0.50
AI_{MOD} = 0.67


Oak 650 Fresh
Bio-Resistant (1808, 41%)
15% carbon loss
H/C Ratio
O/C Ratio
AI_{MOD} = 0.00
AI_{MOD} = 0.50
AI_{MOD} = 0.67


Oak 650 Photo
Bio-Resistant (1651, 37%)
23% carbon loss
O/C Ratio
AI_{MOD} = 0.00
AI_{MOD} = 0.50
AI_{MOD} = 0.67

90

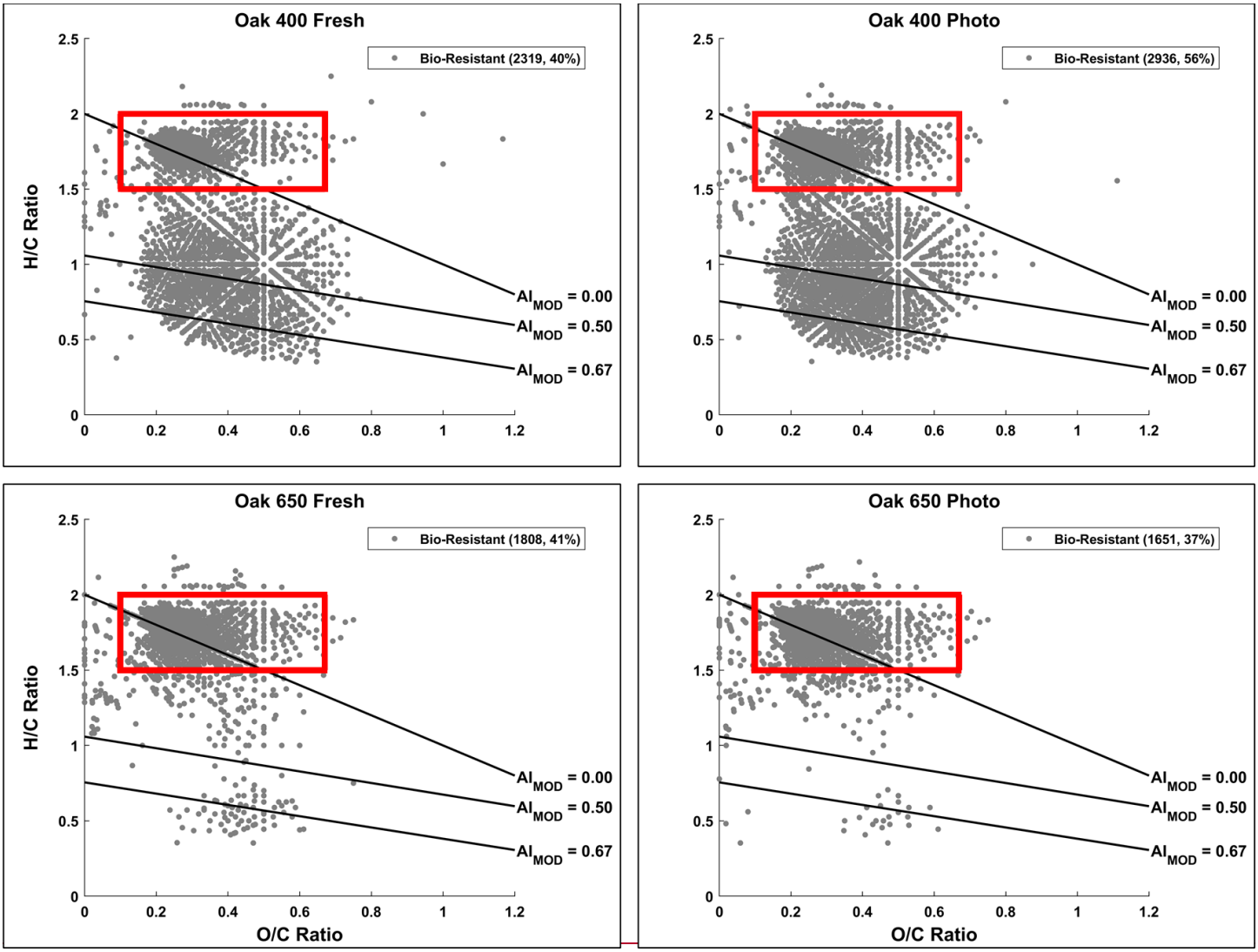
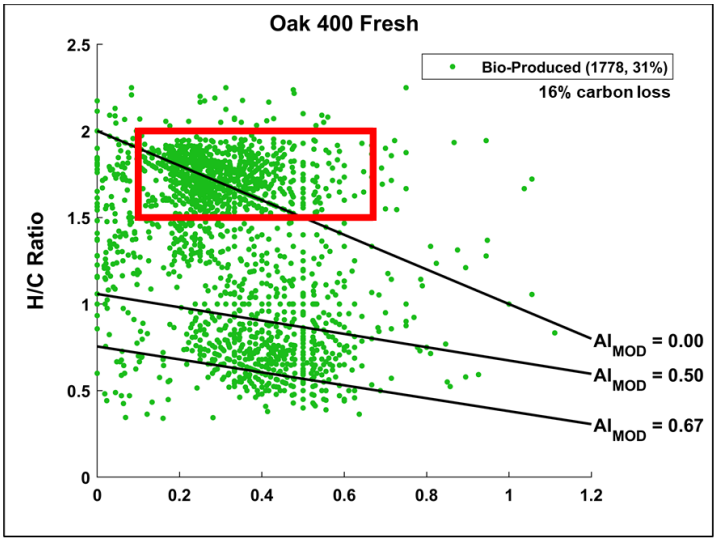
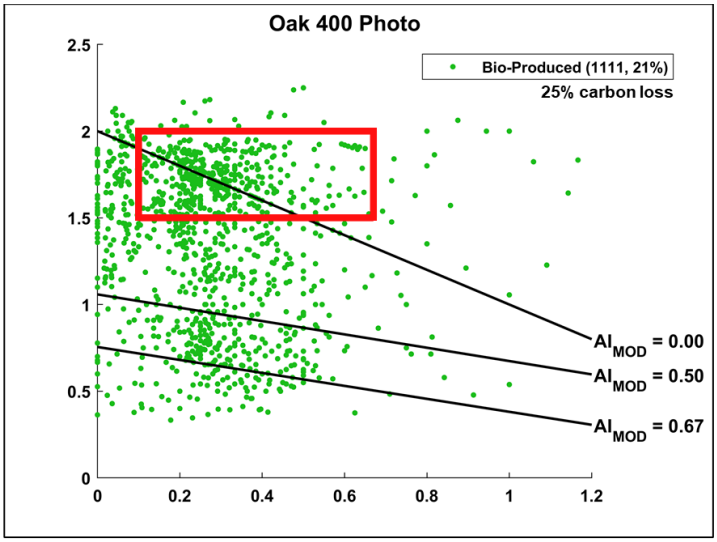
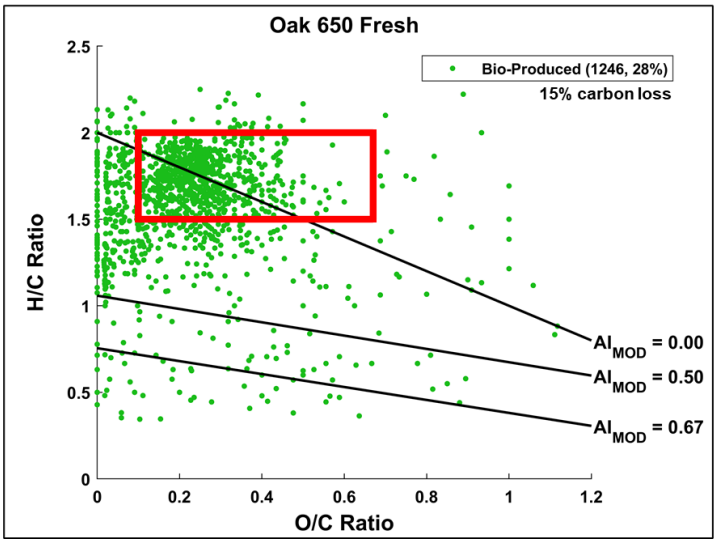


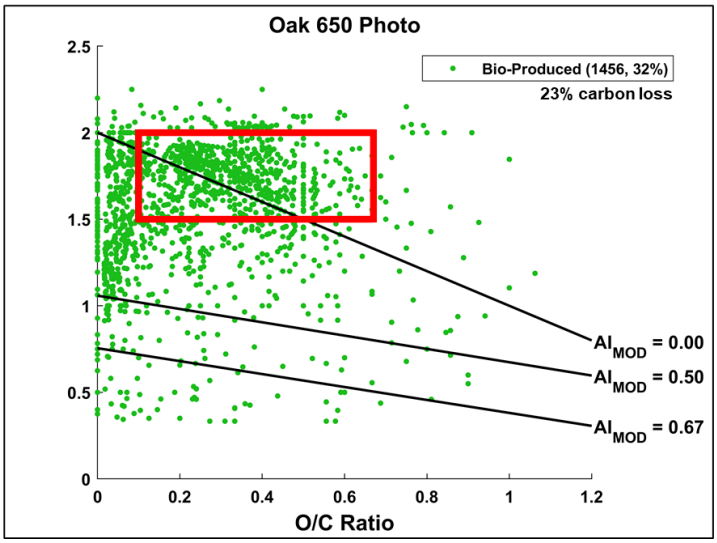
Figure S3. Van Krevelen diagrams of **bio-resistant** formulas identified in the four pyDOM samples using a presence~~/~~-absence approach (Sleighter et al., 2012). The number of formulas and the corresponding percentage (relative to the total number of formulas found in both samples~~in the two samples 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 (AI_{MOD}; Koch and Dittmar, 2006, 2016), and the **red** box indicates the peptide region (valid only for N-containing formulas).

112


Oak 400 Fresh
Bio-Produced (1778, 31%)
16% carbon loss
H/C Ratio
AI_{MOD} = 0.00
AI_{MOD} = 0.50
AI_{MOD} = 0.67


Oak 400 Photo
Bio-Produced (1111, 21%)
25% carbon loss
AI_{MOD} = 0.00
AI_{MOD} = 0.50
AI_{MOD} = 0.67


Oak 650 Fresh
Bio-Produced (1246, 28%)
15% carbon loss
H/C Ratio
O/C Ratio
AI_{MOD} = 0.00
AI_{MOD} = 0.50
AI_{MOD} = 0.67


Oak 650 Photo
Bio-Produced (1456, 32%)
23% carbon loss
O/C Ratio
AI_{MOD} = 0.00
AI_{MOD} = 0.50
AI_{MOD} = 0.67

113

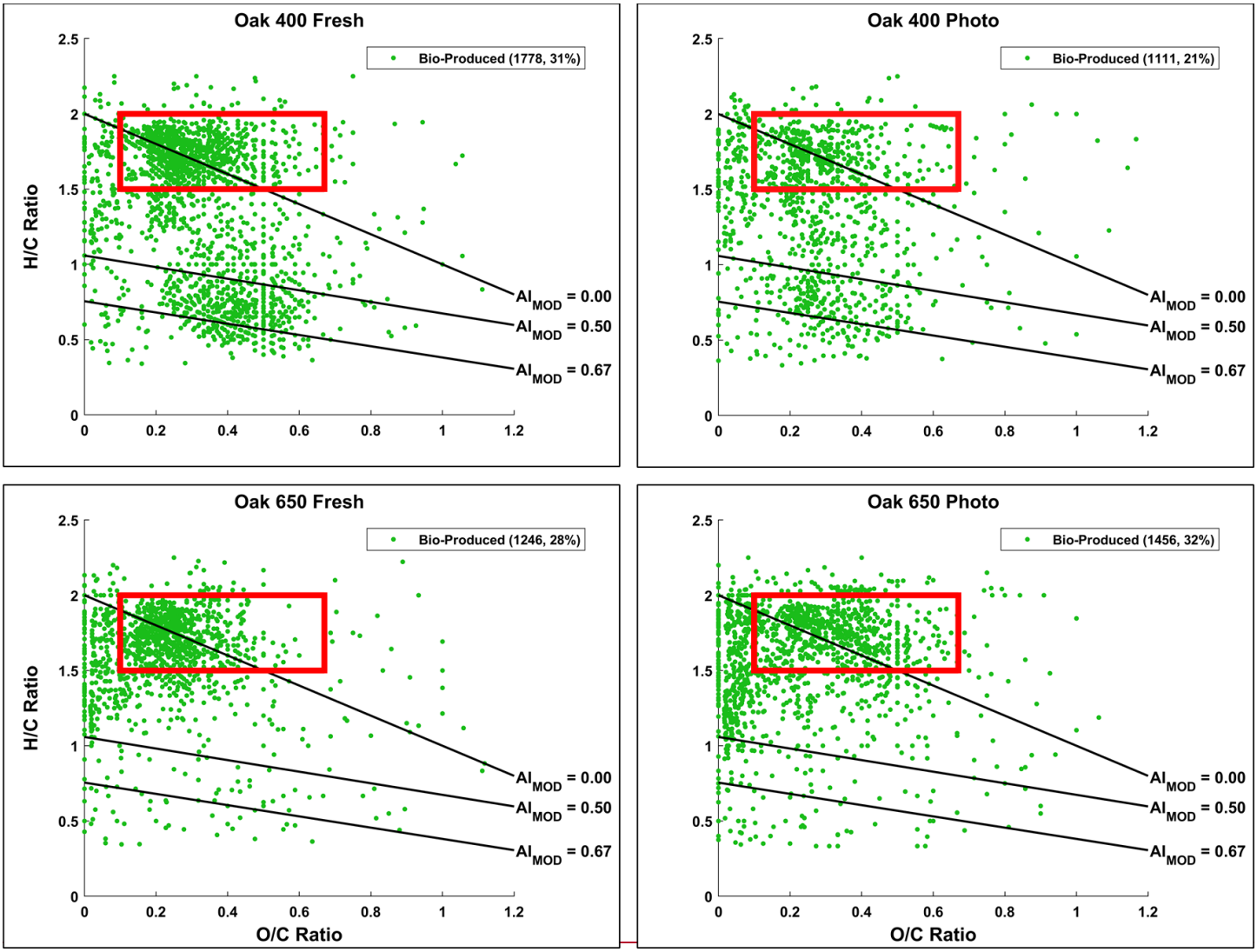
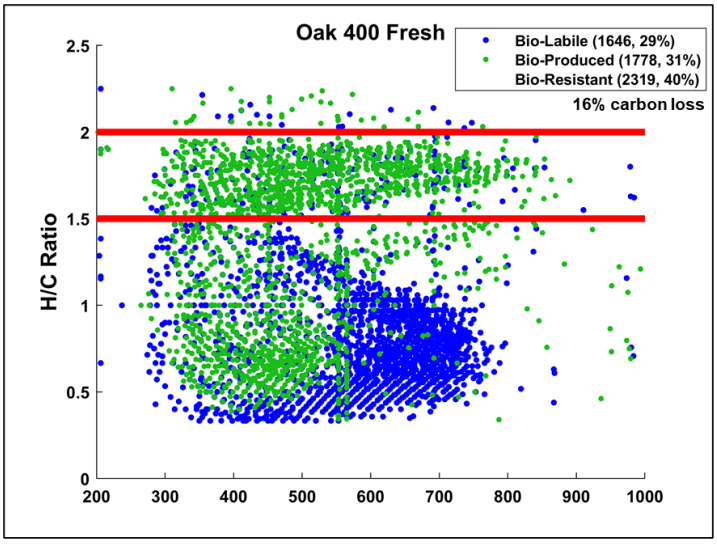
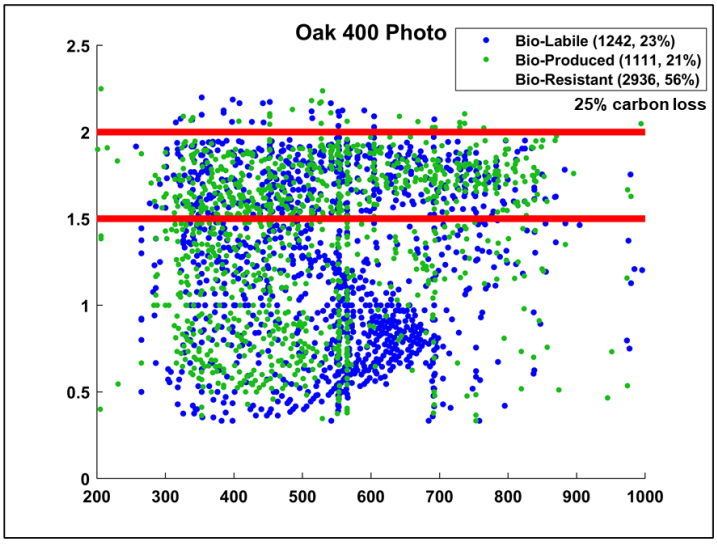
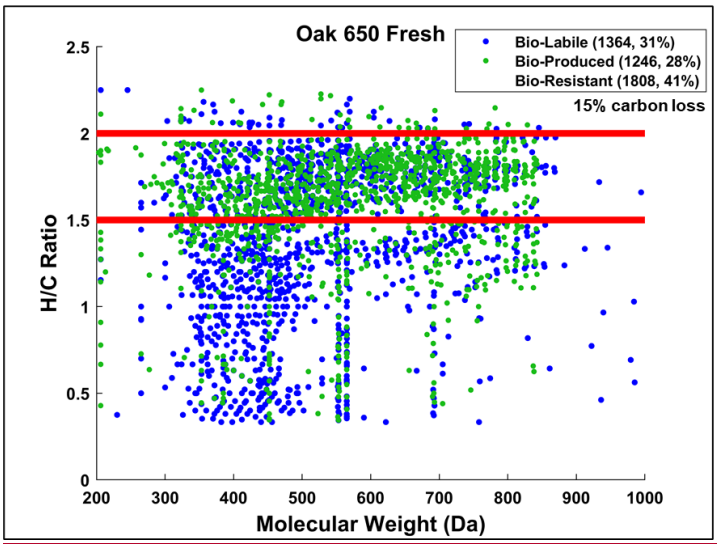


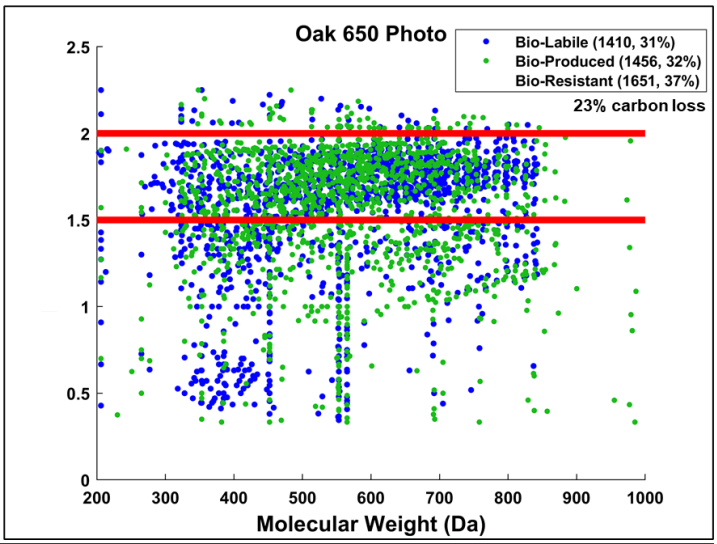
Figure S4. Van Krevelen diagrams of **bio-produced** formulas identified in pyDOM samples using a presence~~/~~-absence approach (Sleighter et al., 2012). The number of formulas and the corresponding percentage (relative to the total number of formulas found in both samples~~in the two samples being 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 (AI_{MOD}; 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~~ molecular ~~Weight~~ weight ~~analysis~~plots


Oak 400 Fresh
Bio-Labile (1646, 29%)
Bio-Produced (1778, 31%)
Bio-Resistant (2319, 40%)
16% carbon loss
H/C Ratio


Oak 400 Photo
Bio-Labile (1242, 23%)
Bio-Produced (1111, 21%)
Bio-Resistant (2936, 56%)
25% carbon loss


Oak 650 Fresh
Bio-Labile (1364, 31%)
Bio-Produced (1246, 28%)
Bio-Resistant (1808, 41%)
15% carbon loss
H/C Ratio
Molecular Weight (Da)


Oak 650 Photo
Bio-Labile (1410, 31%)
Bio-Produced (1456, 32%)
Bio-Resistant (1651, 37%)
23% carbon loss
Molecular Weight (Da)

136

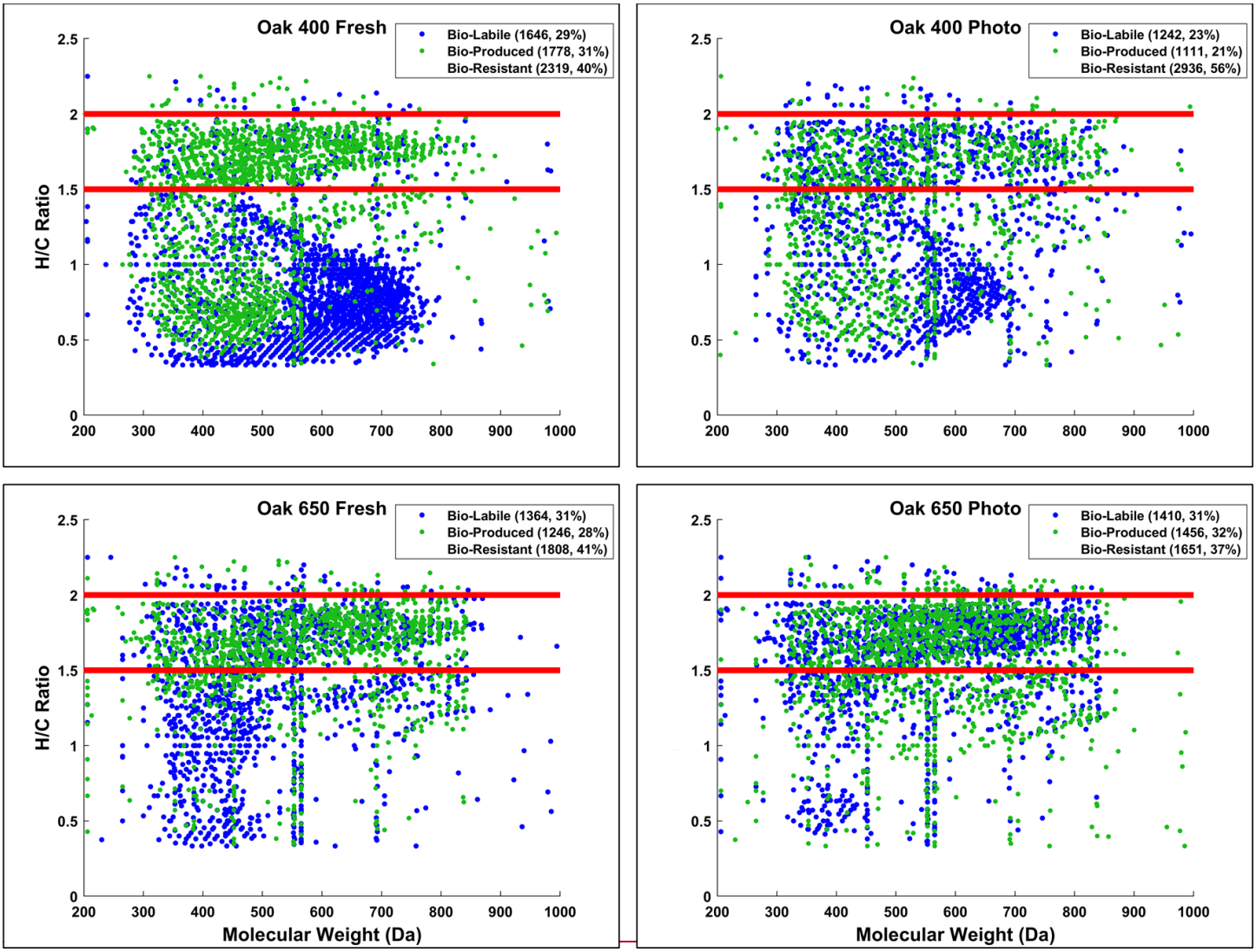
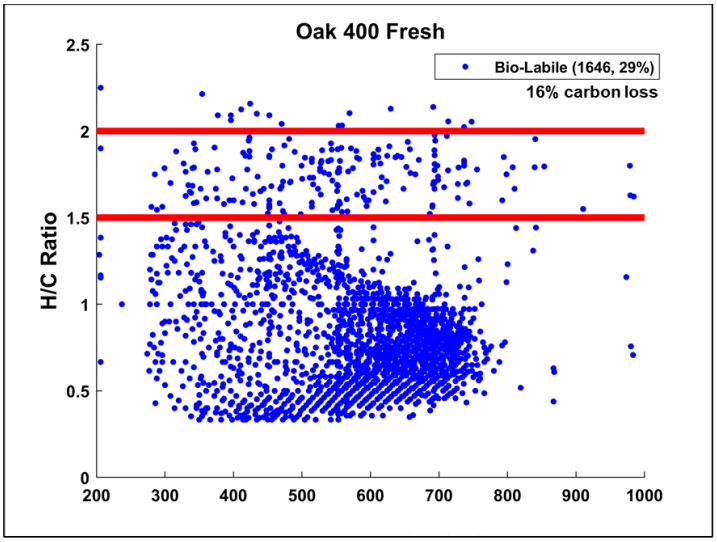
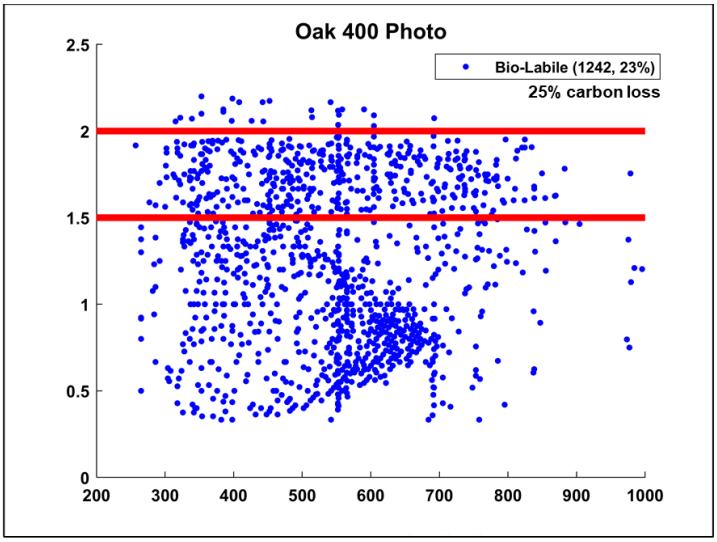
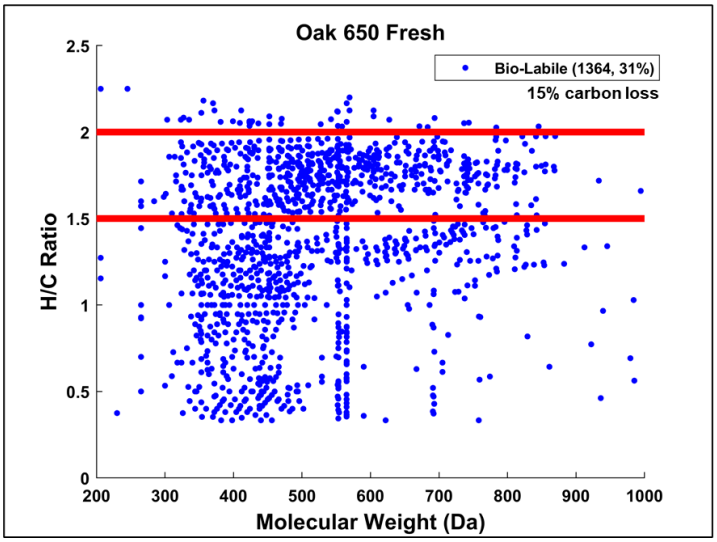
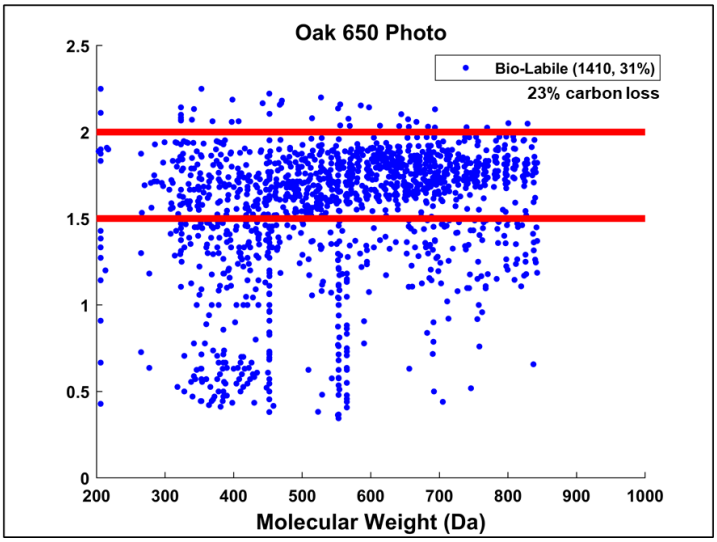


Figure S5. Hydrogen-to-carbon (H/C) ratio versus molecular weight plots of microbially incubated pyDOM leachates. Formulas are classified as **bio-labile** (~~molecular~~ formulas only found in the "killed" control ~~(Fresh or Photo)~~ pyDOM leachates) and **bio-produced** (formulas ~~that are~~ only found in the bio-incubated samples). Formulas that are present in both the "killed" control and bio-incubated samples are operationally classified as bio-resistant and not shown for clarity. These classes are also individually plotted on Figs. S6-8. The number of formulas in~~of~~ each of these pools is shown in the legends (along with their corresponding percentages of the total number of formulas found in both samples). The carbon losses quantified by Bostick et al. (2021) are listed under the legends. The **red** lines indicate where peptide-like formulas would plot.


Oak 400 Fresh
Bio-Labile (1646, 29%)
16% carbon loss
H/C Ratio


Oak 400 Photo
Bio-Labile (1242, 23%)
25% carbon loss


Oak 650 Fresh
Bio-Labile (1364, 31%)
15% carbon loss
H/C Ratio
Molecular Weight (Da)


Oak 650 Photo
Bio-Labile (1410, 31%)
23% carbon loss
Molecular Weight (Da)

146

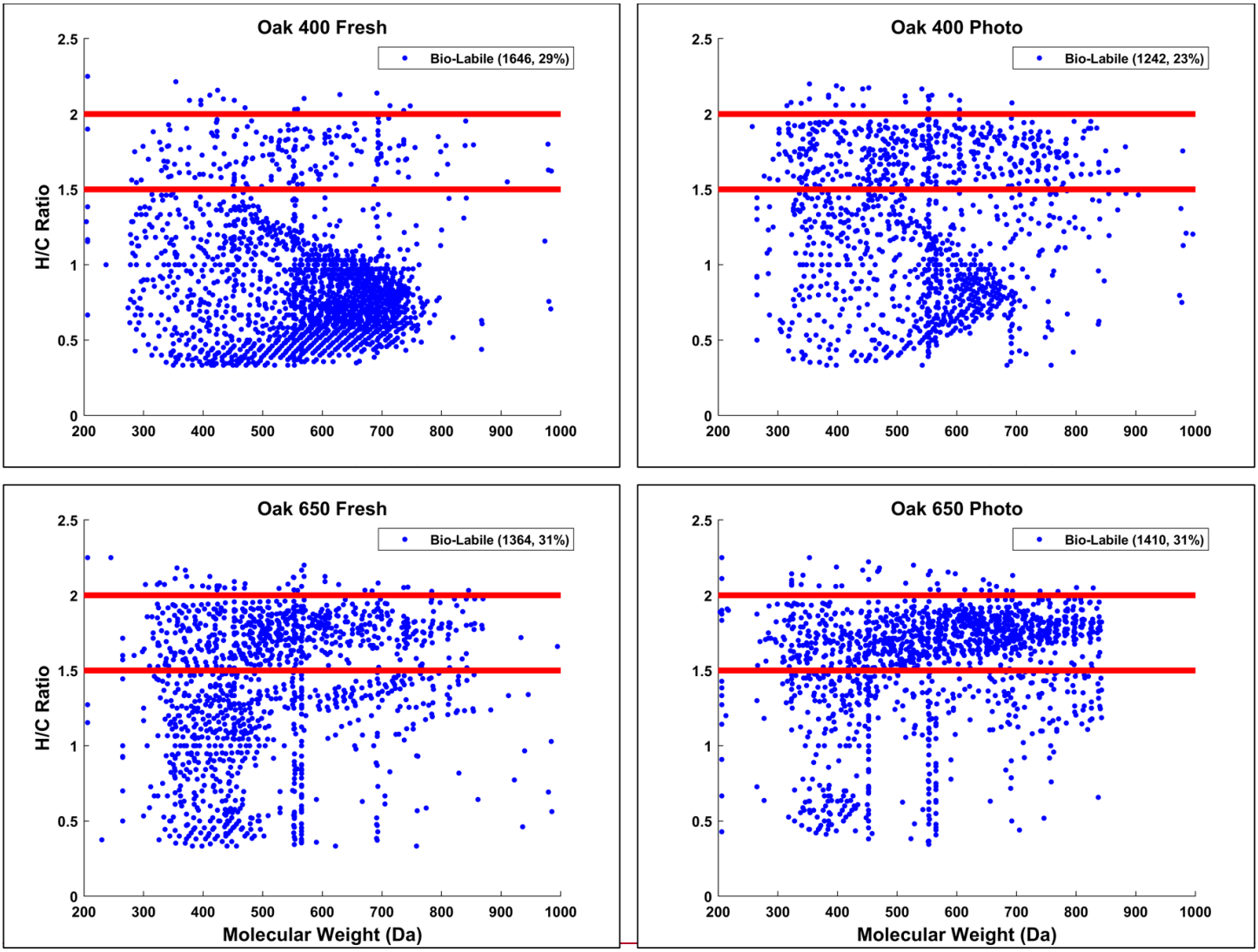
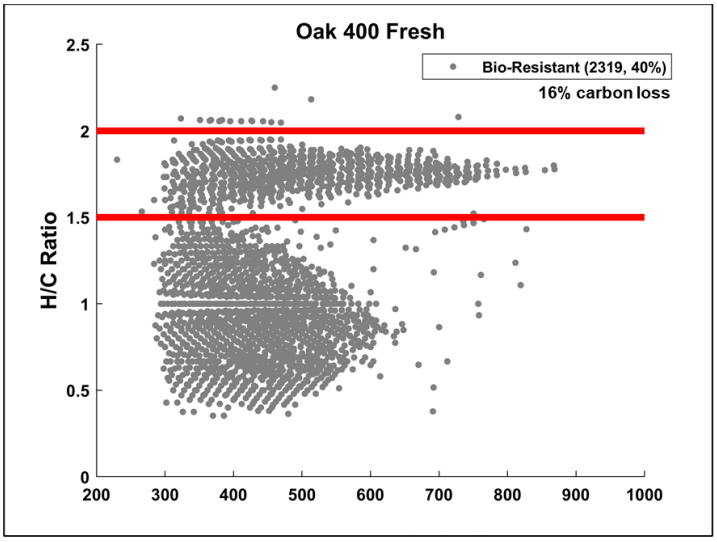
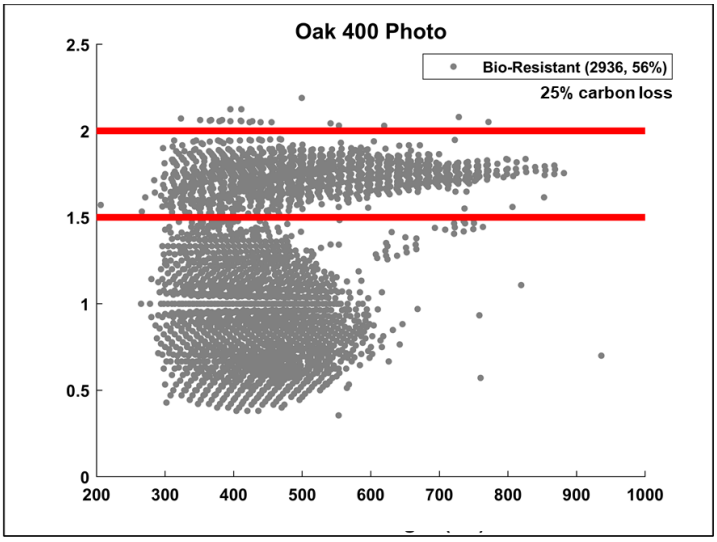
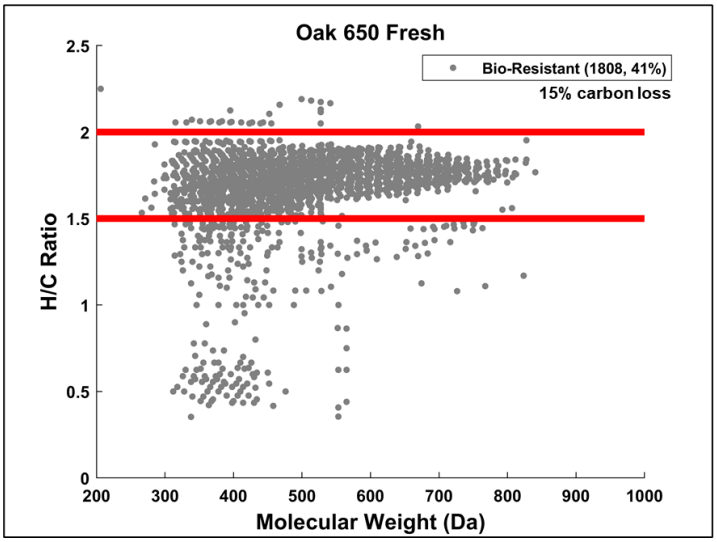
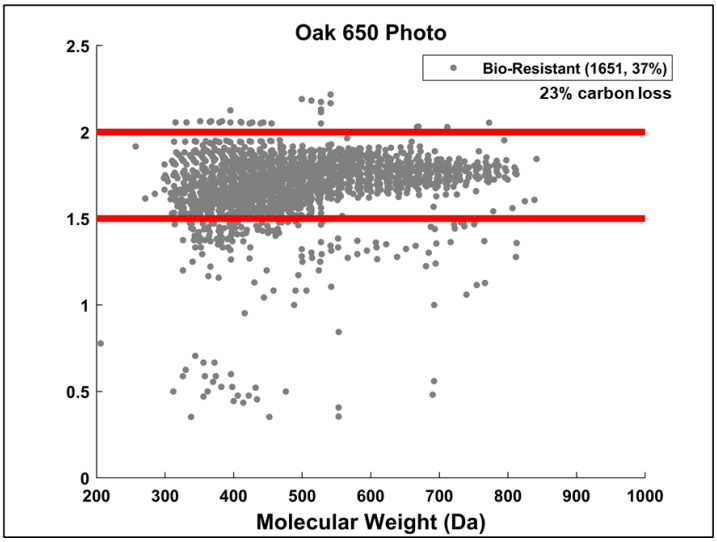


Figure S6. Hydrogen-to-carbon (H/C) ratio versus molecular weight plots of the **bio-labile** formulas. The number of formulas and the corresponding percentage (relative to the total number of formulas found in both samples ~~in the two samples being compared~~) are shown in the legends. The carbon losses quantified by Bostick et al. (2021) are listed under the legends. The **red** lines indicate where peptide-like formulas would plot.


Oak 400 Fresh
Bio-Resistant (2319, 40%)
16% carbon loss
H/C Ratio


Oak 400 Photo
Bio-Resistant (2936, 56%)
25% carbon loss


Oak 650 Fresh
Bio-Resistant (1808, 41%)
15% carbon loss
H/C Ratio
Molecular Weight (Da)


Oak 650 Photo
Bio-Resistant (1651, 37%)
23% carbon loss
Molecular Weight (Da)

152

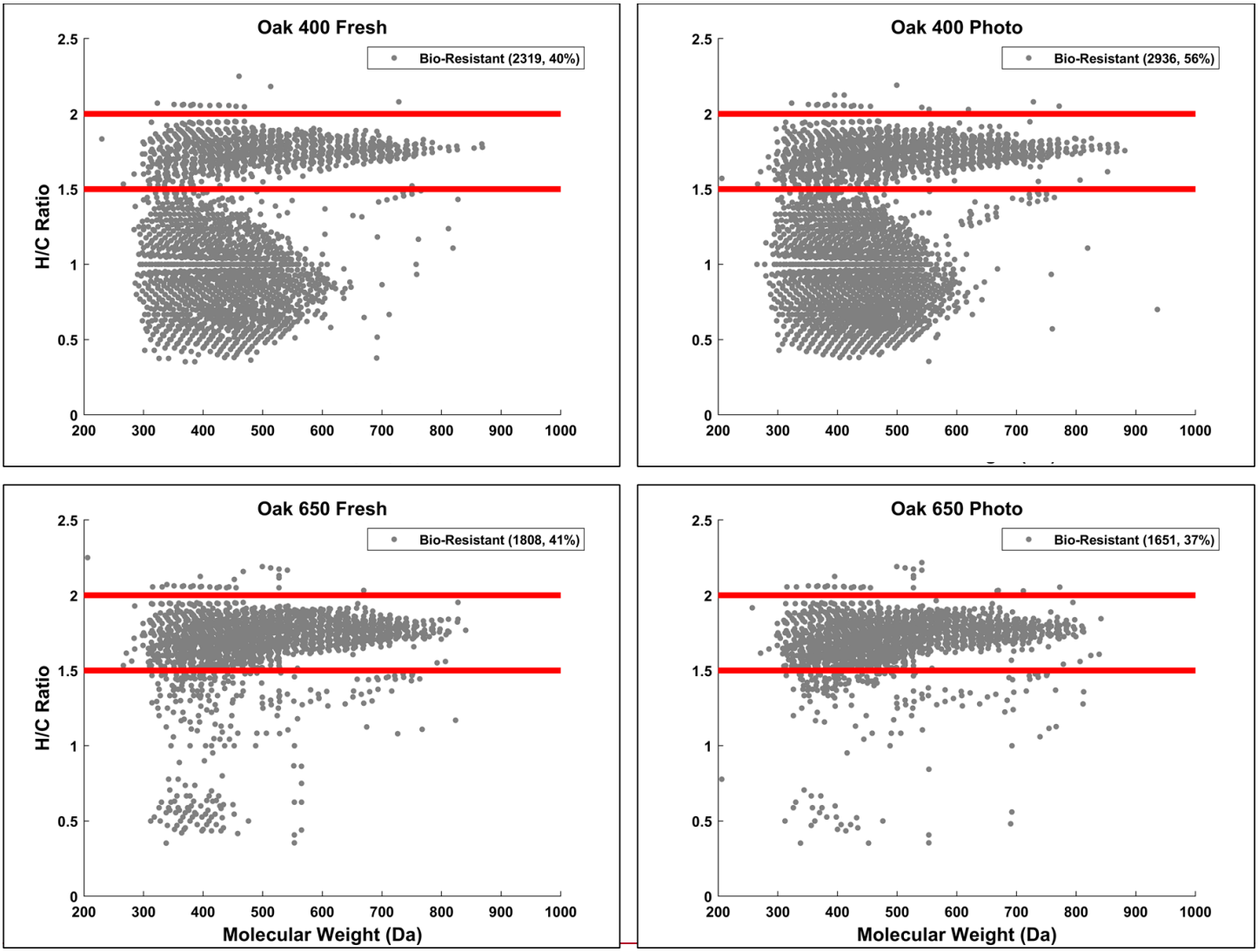
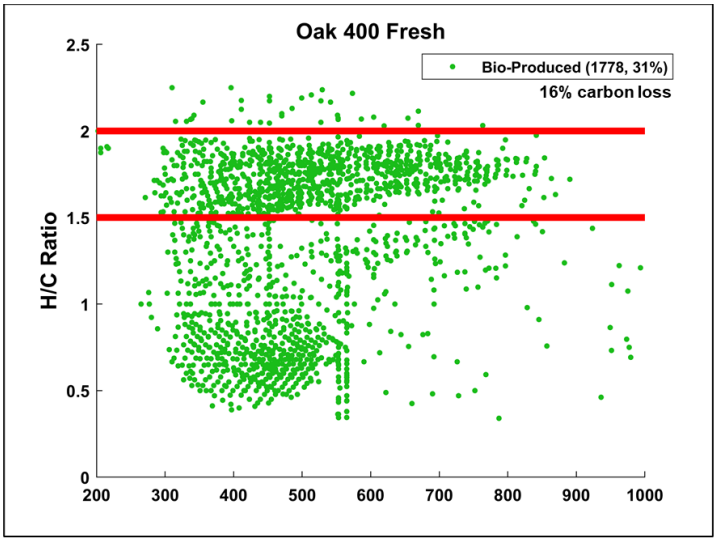
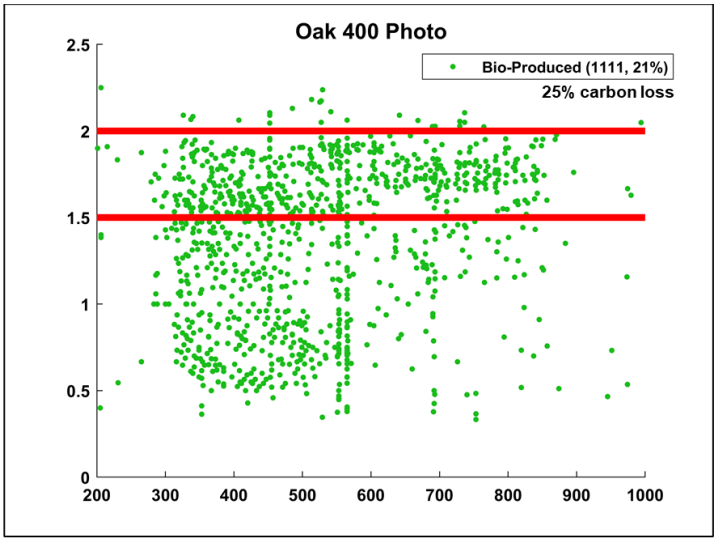
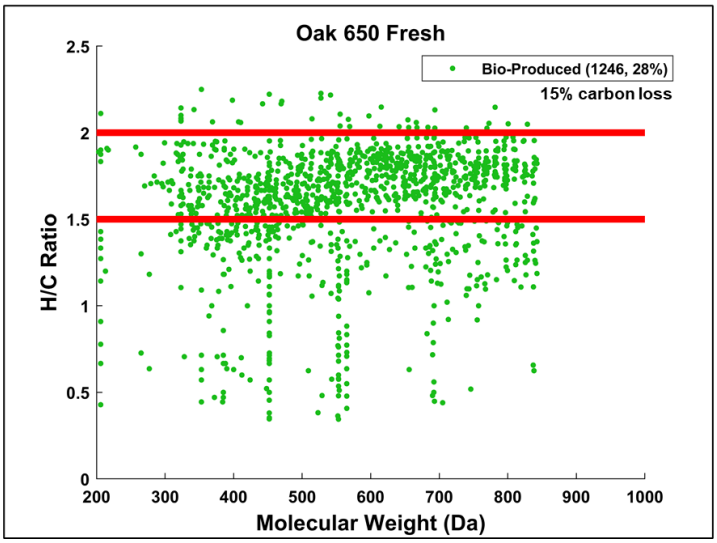
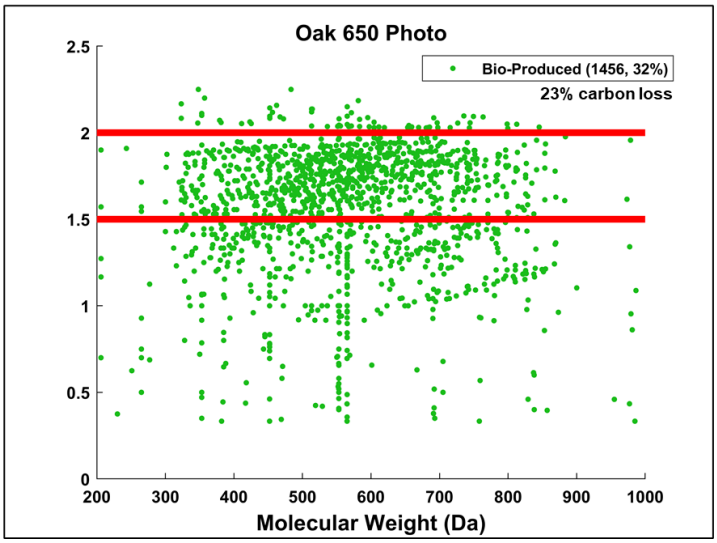


Figure S7. Hydrogen-to-carbon (H/C) ratio versus molecular weight plots of the **bio-resistant** formulas. The number of formulas and the corresponding percentage (relative to the total number of formulas found in both samples~~in the two samples being compared~~) are shown in the legends. The carbon losses quantified by Bostick et al. (2021) are listed under the legends. The **red** lines indicate where peptide-like formulas would plot.


Oak 400 Fresh
Bio-Produced (1778, 31%)
16% carbon loss
H/C Ratio


Oak 400 Photo
Bio-Produced (1111, 21%)
25% carbon loss


Oak 650 Fresh
Bio-Produced (1246, 28%)
15% carbon loss
H/C Ratio
Molecular Weight (Da)


Oak 650 Photo
Bio-Produced (1456, 32%)
23% carbon loss
Molecular Weight (Da)

176

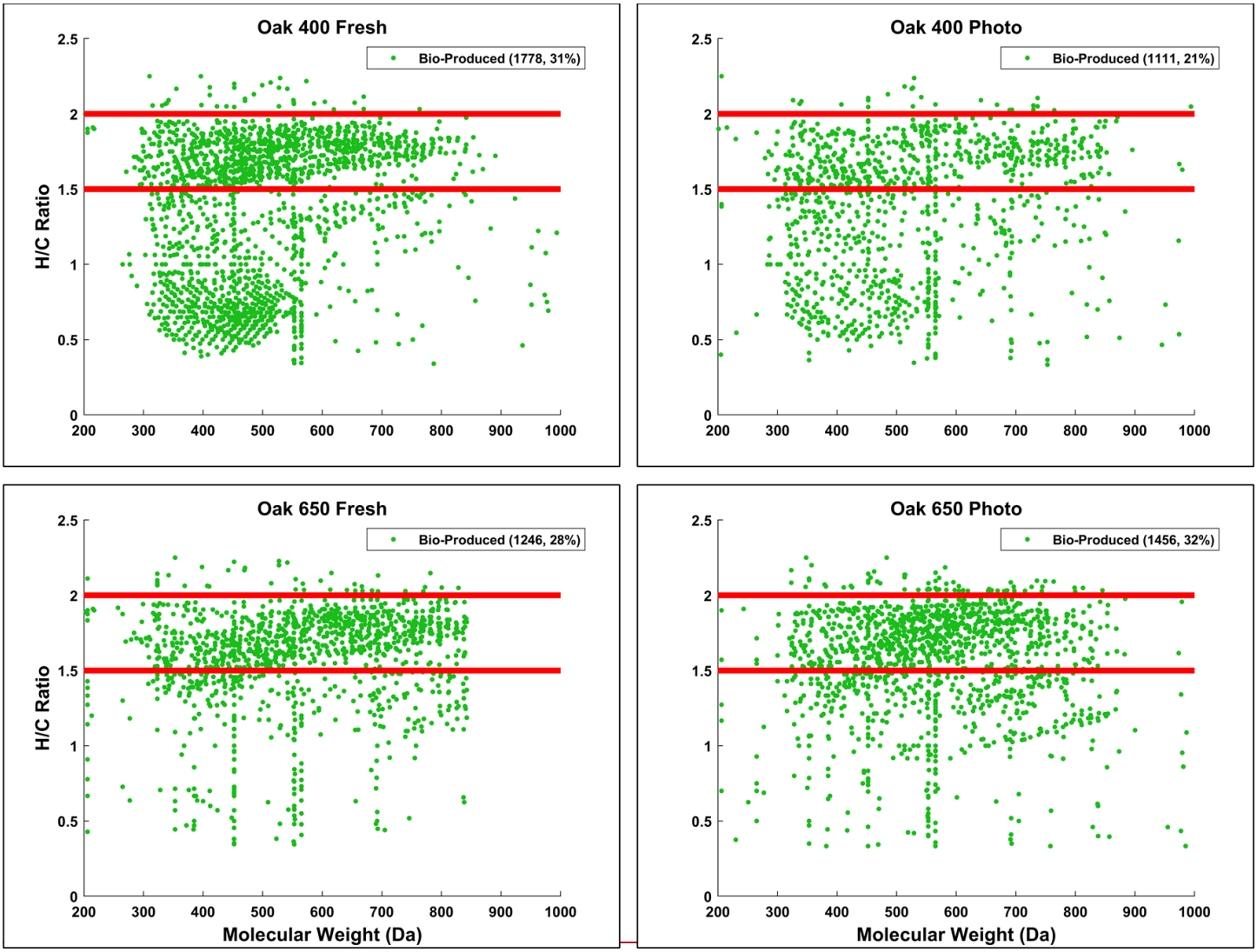


Figure S8. Hydrogen-to-carbon (H/C) ratio versus molecular weight plots of the **bio-produced** formulas. The number of formulas and the corresponding percentage (relative to the total number of formulas found in both samples~~in the two samples being compared~~) are shown in the legends. The carbon losses quantified by Bostick et al. (2021) are listed under the legends. The **red** lines indicate where peptide-like formulas would plot.

Section 4. Evaluation of ~~B~~bio-resistant formulas ~~evaluation~~

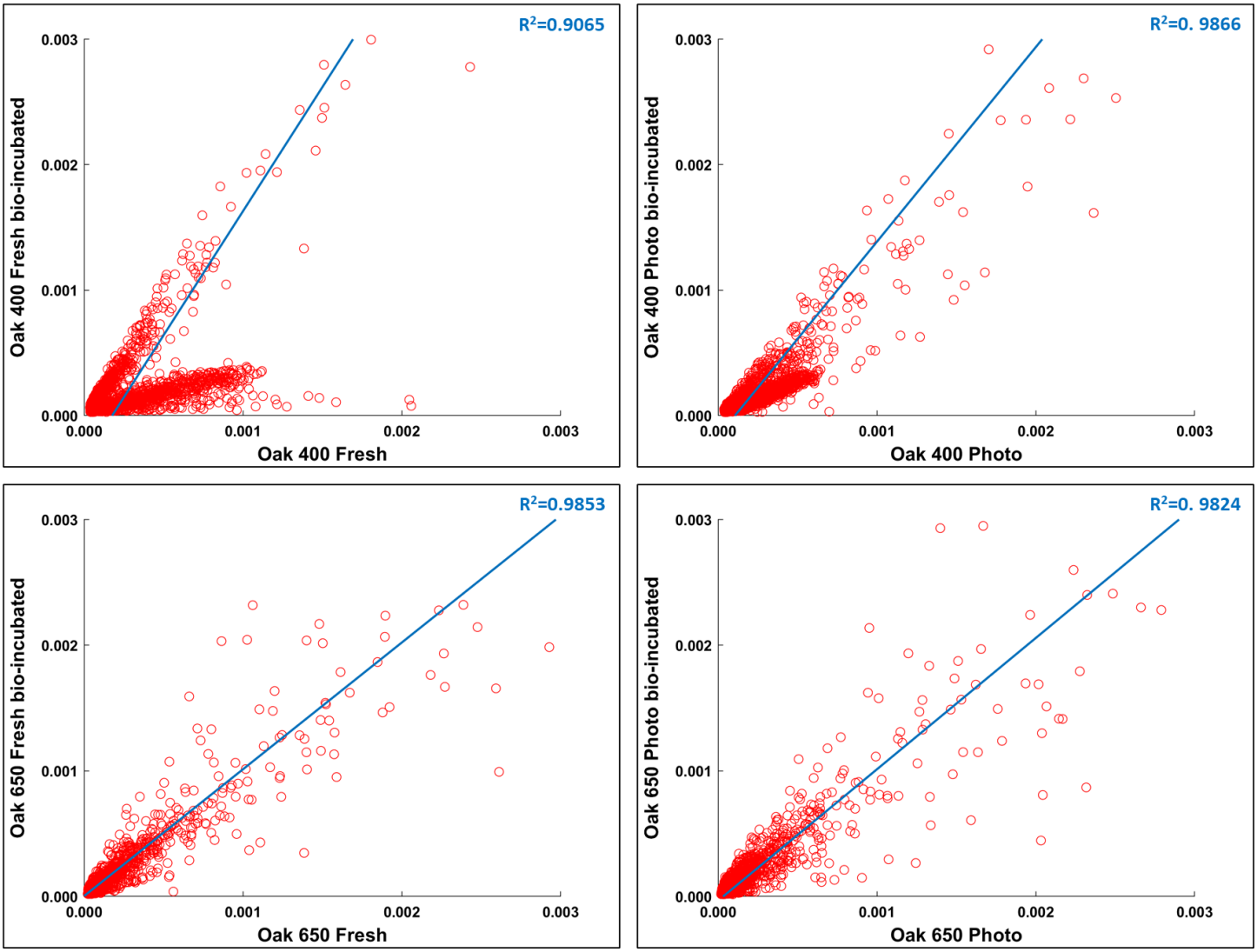


Figure S9. Abundance scatterplots of the bio-resistant formulas following Sleighter et al. (2012). This approach evaluates the similarity in relative abundance of each common formula among the control and its corresponding bio-incubated sample. A high R^2 value indicates a high similarity in the abundance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

Section 5. Comparison of bio-produced formulas with ~~marine~~ marine DOM samples

~~For this analysis, the b~~Bio-produced formulas ~~after~~ from the ~~four~~ pyDOM incubations were combined into one master mass list (total of 4762 formulas). These formulas were searched in environmental ~~previously published molecular d~~data to test ~~whether or not~~if bi~~otic~~ o-incubations of pyDOM produce~~d~~ molecular formulas that have been found across an estuarine transect or in the ocean~~marine-like DOM~~. Bio-produced formulas that were found common to environmental DOM were ~~tested if they were of molecular composition~~ attributed to carboxyl-rich alicyclic molecules (CRAM) ~~based on~~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 ~~marine~~ estuarine transect and oceanic 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), 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.

Sample Name	Number of Formulas	Number of formulas in common with all bio-produced formulas of pyDOM	
Dismal Swamp [0], C18[a]	1752~~1752~~	223 (13 %)4 (~0%)	CRAM: 123 (55%)
Great Bridge [11], C18[a]	1727~~1727~~	292 (17 %)6 (~0%)	CRAM: 148 (51%)
Town Point [20], C18[a]	1303~~1303~~	228 (18 %)4 (~0%)	CRAM: 102 (45%)
~~C~~Chesapeake Bay ~~C~~Bridge [22], C18[a]	1079~~1079~~	193 (18 %)4 (~0%)	CRAM: 81 (42%)
Off Shore Coast [32], C18[a]	1189~~1189~~	212 (18 %)4 (~0%)	CRAM: 88 (42%)
N. Atlantic Ocean surface water (DOM411), PPL[b]	2402~~2402~~	227 (10 %)3 (~0%)	CRAM: 159 (70%)
N. Atlantic Ocean surface water ~~DOM~~(DOM412), PPL[b]	3524~~3524~~	289 (8 %)6 (~0%)	CRAM: 192 (66%)
N. Atlantic Ocean surface water ~~DOM~~(DOM417), PPL[b]	3312~~3312~~	263 (8 %)3 (~0%)	CRAM: 181 (69%)
N. Pacific Ocean surface water (DOM 1), RO/ED[c,d]	1918~~1697~~	261 (14 %)~~249~~ (~5%)	CRAM: 155 (59%)
N. Pacific Ocean surface water (DOM 1-r~~ep~~), RO/ED[c,d]	1950~~1756~~	258 (13 %)~~272~~ (~6%)	CRAM: 152 (59%)
N. Atlantic Ocean abyssal water (DOM 2), RO/ED[c,d]	1697~~1918~~	~~223~~ 284 (~~5~~17 %)	CRAM: 154 (54%)
N. Atlantic Ocean abyssal water (DOM 2-r~~ep~~), RO/ED[c,d]	1756~~1950~~	308 (18 %)~~219~~ (~5%)	CRAM: 167 (54%)
Coastal N. Pacific Ocean water (DOM 3), PPL[d]	2226~~2226~~	265 (12 %)~~223~~ (~5%)	CRAM: 161 (61%)
Coastal N. Pacific Ocean water (DOM 3 rep), PPL[d]	2256~~2256~~	278 (12 %)~~235~~ (~5%)	CRAM: 168 (60%)
Coastal N. Pacific Ocean water (DOM 4), PPL[d]	2325~~2325~~	287 (12 %)~~246~~ (~5%)	CRAM: 169 (59%)

Coastal N. Pacific Ocean water (DOM 4 rep), PPL[d]	24292429	288 (12 %)244 (~5%)	CRAM: 177 (61%)

[a]Estuarine transect data from Sleighter and Hatcher (2008). Only the Off Shore Coast sample is considered oceanic.
[b]Unpublished data from samples obtained during the WACS-2 cruise (*R/V Knorr*) as part of the Western Atlantic Climate Study (WACS).
[c]Chen et al. (2014)
[d]Sleighter et al. (2012)

The four different pyDOM samples were also individually compared to all oceanic 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.

Section 6. Analysis of ~~Variance~~ variance (ANOVA) of bio-produced peptide-like ~~organic matter by the pyDOM samples, as well as by the sucrose reference sample.~~formulas

Table S2. Molecular metrics of peptide-like bio-produced formulas (N-containing, $1.5 \leq H/C \leq 2.0$, $0.1 \leq O/C \leq 0.67$) 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 ~~(evaluation done by ANOVA followed by Scheffé's post-hoc test)~~.

	Oak 400 Fresh	Oak 400 Photo	Oak 650 Fresh	Oak 650 Photo	Sucrose
Number of bio-produced formulas	1778	1111	1246	1456	1339
Number of peptide-like	541 (30%)	261 (23%)	497 (40%)	314 (22%)	160 (12%)

bio-produced formulas					
Number of identified oligopeptides	14	5	11	18	2
C number	**28.5 ± 7.6**	30.9 ± 10.9	30.7 ± 7.6	30.3 ± 8.7	31.7 ± 9.6
H number	**49.8 ± 14.4**	54 ± 20.6	53.7 ± 14.8	54 ± 16.5	55.4 ± 18.5
O number	7.8 ± 2.6	7.8 ± 3.2	7.8 ± 2.9	**9.0 ± 2.8**	7.9 ± 3.1
N number	2.4 ± 1.1	**2.8 ± 1.3**	2.5 ± 1.2	2.4 ± 1.2	2.4 ± 1.3
O/C ratio	**0.28 ± 0.08**	0.26 ± 0.09	0.25 ± 0.08	**0.31 ± 0.10**	0.25 ± 0.08
H/C ratio	1.74 ± 0.12	1.74 ± 0.13	1.74 ± 0.13	**1.78 ± 0.16**	1.74 ± 0.14
N/C ratio	0.085 ± 0.037	**0.094 ± 0.045**	0.082 ± 0.038	0.083 ± 0.045	0.078 ± 0.042
H/N ratio	24.8 ± 11.4	23.5 ± 13.4	26 ± 13.2	**28.6 ± 16.7**	**29.4 ± 16**
O/N ratio	4.0 ± 2.2	**3.5 ± 2.2**	3.8 ± 2.5	**5.1 ± 3.5**	4.3 ± 2.7
MW[a]	**550 ± 140**	589 ± 188	582 ± 147	596 ± 143	597 ± 172
DBE[b]	5.81 ± 1.78	6.28 ± 2.17	6.13 ± 2.06	**5.51 ± 2.59**	6.2 ± 2.33
DBE/C[c]	0.211 ± 0.065	0.215 ± 0.071	0.206 ± 0.069	**0.189 ± 0.083**	0.203 ± 0.071
DBE-O[d]	-2.27 ± 2.75	-1.75 ± 3.52	-1.90 ± 3.55	**-3.82 ± 4.26**	-1.86 ± 3.65
AI_{MOD}[e]	0.077 ± 0.05	0.090 ± 0.052	0.083 ± 0.049	**0.089 ± 0.057**	**0.116 ± 0.049**
NOSC[f]	-0.929 ± 0.239	-0.933 ± 0.259	**-0.984 ± 0.227**	-0.903 ± 0.269	**-1.002 ± 0.218**

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

The proteinaceous formulas in the four samples were evaluated using one-way ANOVA to ~~extract~~ assess the variability in their composition. While the peptide-like formulas seemed similar when plotted in the van Krevelen space (Figs. 1 and S3). Averages of molecular parameters were derived from the 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_{MOD}) and nominal oxidation state of carbon (NOSC). ~~Averages of molecular parameters were derived from the 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_{MOD}) and nominal oxidation state of carbon (NOSC). While the peptide-like formulas seem similar when plotted in the vK space (Figs. 1 and S3), significant differences ($p < 0.05$) in the means of all molecular parameters were observed.~~ 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 ~~the vast diversity of~~that the bio-produced peptide-like molecules ~~after these five incubations~~were of vast diversity among the different incubations.

Section 7. Oligopeptide Sequences

Table S3. Oligopeptide sequences ~~found in~~consistent with the bio-produced formulas of each pyDOM sample.

Sample	Measured m/z	Amino Acid combination[#]	Molecular weight (Da)	Molecular Formula
Oak 400 Fresh	201.1246	AL	202.1317	$C_9H_{18}O_3N_2$
Oak 400 Fresh	356.2192	OLL	357.2264	$C_{17}H_{31}O_5N_3$
Oak 400 Fresh	455.2874	OLLV	456.2948	$C_{22}H_{40}O_6N_4$

Oak 400 Fresh	512.3457	ALLVV	513.3526	$C_{25}H_{47}O_6N_5$
Oak 400 Fresh	512.3457	GLLLV	513.3526	$C_{25}H_{47}O_6N_5$
Oak 400 Fresh	512.3457	VVVVV	513.3526	$C_{25}H_{47}O_6N_5$
Oak 400 Fresh	514.3251	ALLLS	515.3319	$C_{24}H_{45}O_7N_5$
Oak 400 Fresh	514.3251	ALLTV	515.3319	$C_{24}H_{45}O_7N_5$
Oak 400 Fresh	514.3251	GLLLT	515.3319	$C_{24}H_{45}O_7N_5$
Oak 400 Fresh	514.3251	LSVVV	515.3319	$C_{24}H_{45}O_7N_5$
Oak 400 Fresh	514.3251	TVVVV	515.3319	$C_{24}H_{45}O_7N_5$
Oak 400 Fresh	526.3607	ALLLV	527.3683	$C_{26}H_{49}O_6N_5$
Oak 400 Fresh	526.3607	GLLLL	527.3683	$C_{26}H_{49}O_6N_5$
Oak 400 Fresh	526.3607	LVVVV	527.3683	$C_{26}H_{49}O_6N_5$

Oak 400 Photo	341.2195	LPX	342.2267	$C_{16}H_{30}O_4N_4$
Oak 400 Photo	341.2195	KPV	342.2267	$C_{16}H_{30}O_4N_4$
Oak 400 Photo	350.1836	HPV	351.1907	$C_{16}H_{25}O_4N_5$
Oak 400 Photo	528.3188	LLWV	529.3264	$C_{28}H_{43}O_5N_5$
Oak 400 Photo	552.3768	LLLPV	553.3839	$C_{28}H_{51}O_6N_5$

Oak 650 Fresh	498.3293	AALLL	499.3370	$C_{24}H_{45}O_6N_5$
Oak 650 Fresh	498.3293	ALVVV	499.3370	$C_{24}H_{45}O_6N_5$
Oak 650 Fresh	498.3293	GLLVV	499.3370	$C_{24}H_{45}O_6N_5$
Oak 650 Fresh	512.3455	ALLVV	513.3526	$C_{25}H_{47}O_6N_5$
Oak 650 Fresh	512.3455	GLLLV	513.3526	$C_{25}H_{47}O_6N_5$
Oak 650 Fresh	512.3455	VVVVV	513.3526	$C_{25}H_{47}O_6N_5$
Oak 650 Fresh	552.3042	DLLPP	553.3112	$C_{26}H_{43}O_8N_5$
Oak 650 Fresh	552.3042	ELPPV	553.3112	$C_{26}H_{43}O_8N_5$
Oak 650 Fresh	552.3042	OOLPV	553.3112	$C_{26}H_{43}O_8N_5$
Oak 650 Fresh	552.3042	OLUVV	553.3112	$C_{26}H_{43}O_8N_5$
Oak 650 Fresh	552.3042	LLPUT	553.3112	$C_{26}H_{43}O_8N_5$

Oak 650 Photo	242.1508	KP	243.1583	$C_{11}H_{21}O_3N_3$
Oak 650 Photo	342.2034	OLV	343.2107	$C_{16}H_{29}O_5N_3$
Oak 650 Photo	356.2190	OLL	357.2264	$C_{17}H_{31}O_5N_3$
Oak 650 Photo	552.2676	ALSTY	553.2748	$C_{25}H_{39}O_9N_5$
Oak 650 Photo	552.2676	ATTYV	553.2748	$C_{25}H_{39}O_9N_5$
Oak 650 Photo	552.2676	DOLPP	553.2748	$C_{25}H_{39}O_9N_5$
Oak 650 Photo	552.2676	DLPUV	553.2748	$C_{25}H_{39}O_9N_5$
Oak 650 Photo	552.2676	EOPPV	553.2748	$C_{25}H_{39}O_9N_5$
Oak 650 Photo	552.2676	EPUVV	553.2748	$C_{25}H_{39}O_9N_5$
Oak 650 Photo	552.2676	GLTTY	553.2748	$C_{25}H_{39}O_9N_5$
Oak 650 Photo	552.2676	OOOPV	553.2748	$C_{25}H_{39}O_9N_5$
Oak 650 Photo	552.2676	OOUVV	553.2748	$C_{25}H_{39}O_9N_5$
Oak 650 Photo	552.2676	OLPUT	553.2748	$C_{25}H_{39}O_9N_5$
Oak 650 Photo	552.2676	LLUUS	553.2748	$C_{25}H_{39}O_9N_5$

Oak 650 Photo	552.2676	LFSST	553.2748	$C_{25}H_{39}O_9N_5$
Oak 650 Photo	552.2676	LUUTV	553.2748	$C_{25}H_{39}O_9N_5$
Oak 650 Photo	552.2676	FSTTV	553.2748	$C_{25}H_{39}O_9N_5$
Oak 650 Photo	552.2676	SSYVV	553.2748	$C_{25}H_{39}O_9N_5$

Sucrose	340.1880	OLP	341.1951	$C_{16}H_{27}O_5N_3$
Sucrose	340.1880	LUV	341.1951	$C_{16}H_{27}O_5N_3$

[#]Combinations can be of any order

Section 8. Kendrick ~~Mass~~ mass ~~Defect~~ defect ~~Analysis Plots~~ plots of bio-produced formulas

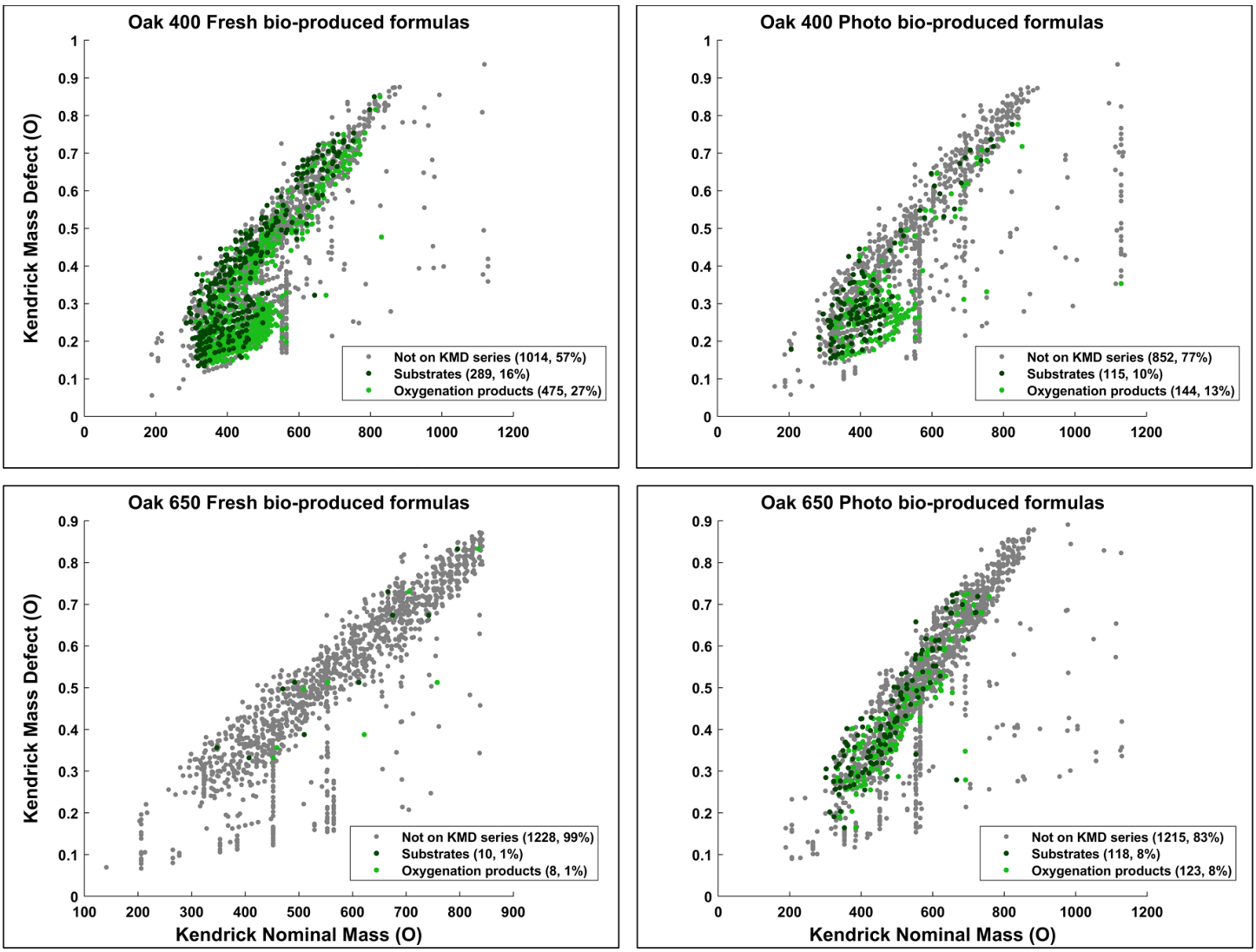


Figure S10. Kendrick ~~Mass~~ mass ~~Defect~~ defect (KMD) versus Kendrick ~~Nominal~~ nominal ~~Mass~~ mass plots for the ~~Oxygen~~ oxygen (O) series within the bio-produced formulas of the four pyDOM samples. Formulas not part of the O KMD series are colored in **gray**. Formulas in **dark green** 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).

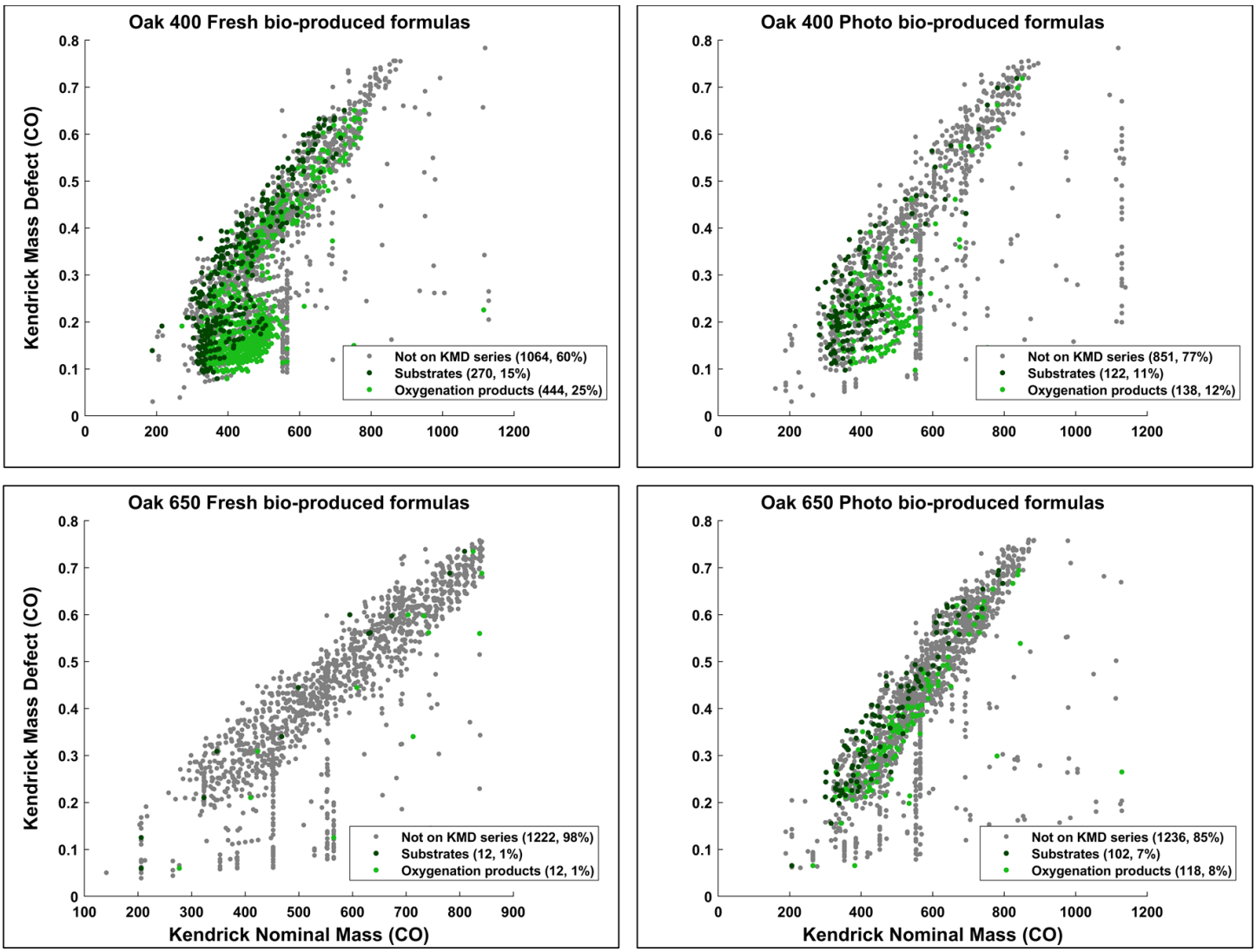


Figure S11. Kendrick ~~Mass~~ mass ~~Defect~~ defect (KMD) versus Kendrick ~~Nominal~~ nominal ~~Mass~~ mass plots for the Carbonyl (CO) series within the bio-produced formulas of the four pyDOM samples. Formulas not part of the CO KMD series are colored in **gray**. Formulas in **dark green** are substrates 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).

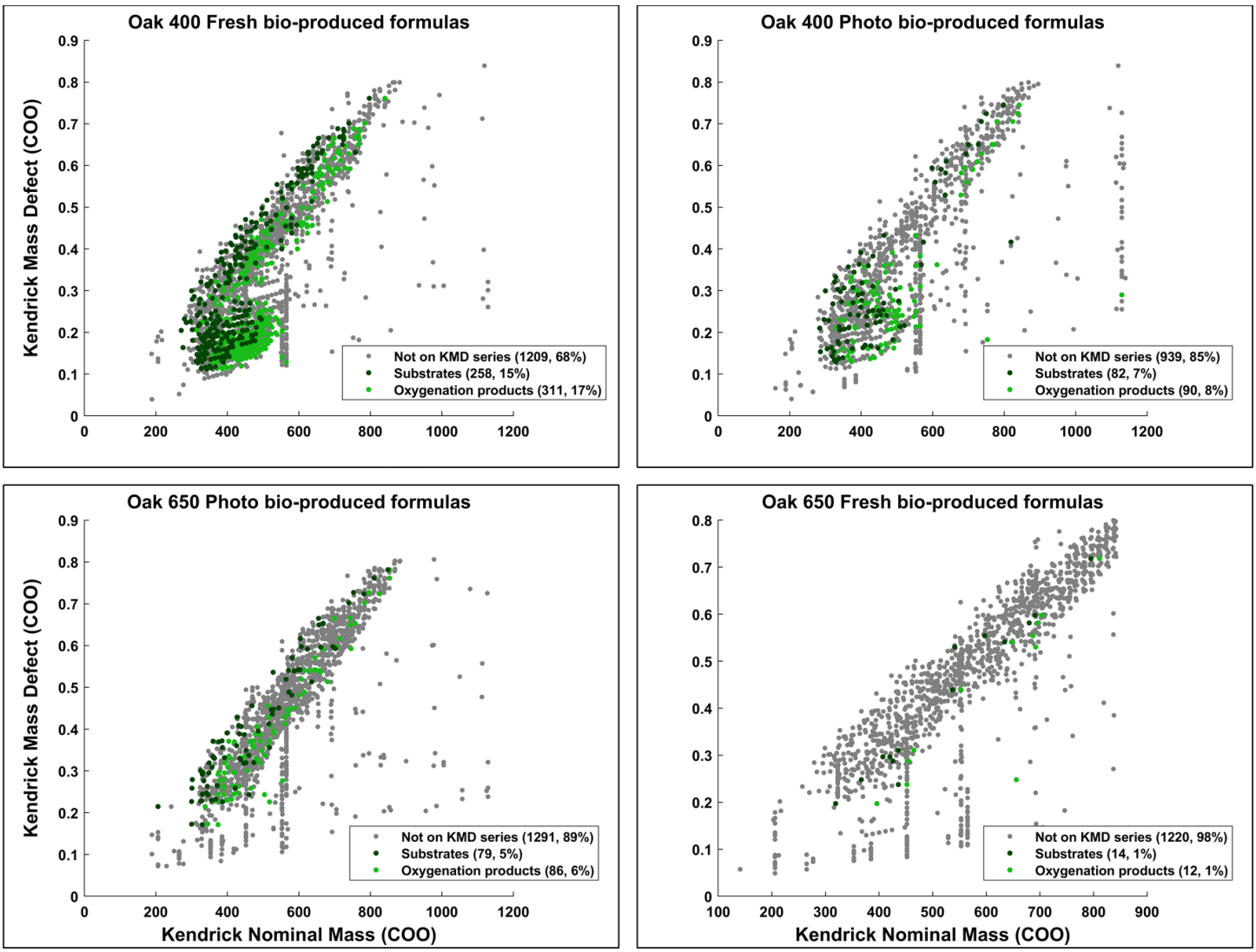


Figure S12. Kendrick ~~Mass~~ mass ~~Defect~~ defect (KMD) versus Kendrick ~~Nominal~~ nominal ~~Mass~~ mass plots for the Carboxyl (COO) series within the bio-produced formulas of the four pyDOM samples. Formulas not part of the COO KMD series are colored in **gray**. Formulas in **dark green** are 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).

Section 9. Correlation analysis of molecular diversity and NMR data

Table S4. Data used for ~~the correlation analysis~~Pearson correlations between molecular diversity (~~as determined by~~ FT-ICR-MS molecular formulas) and 1D NMR (Bostick et al., ~~2020~~2021). ~~Coefficients of determination (~~R^2 and p-values~~)~~ are listed for each functional group in the corresponding color.

	Oak 400 Fresh	Oak 400 Photo	Oak 650 Fresh	Oak 650 Photo
Number of bio-labile formulas	1646	1242	1364	1410
Number of bio-produced formulas	1778	1111	1246	1456
Aldehyde (O=CH) $R^2 = 0.1263$, $R^2 = 0.2374$ $p = 0.6448$, $p = 0.5130$	3.18%	4.52%	10.99%	4.24%
Aryl $R^2 = 0.0094$, $R^2 = 0.0668$ $p = 0.9031$, $p = 0.7418$	9.87%	8.47%	20.65%	7.54%
Olefinic (C=C) $R^2 = 0.9472$, $R^2 = 0.9978$ $p = 0.0267$, $p = 0.0011$	7.64%	15.60%	14.31%	11.41%
HC-O-R $R^2 = 0.4217$, $R^2 = 0.3385$ $p = 0.3509$, $p = 0.4184$	6.75%	23.64%	4.57%	9.41%
HC-C=Y $R^2 = 0.0201$, $R^2 = 0.0511$ $p = 0.8590$, $p = 0.7746$	12.33%	13.14%	4.49%	9.13%
HC-C-C-X $R^2 = 0.4639$, $R^2 = 0.3968$ $p = 0.3201$, $p = 0.3930$	3.98%	5.99%	6.52%	7.38%
Methylene (CH_2) $R2 = 0.1287$, $R^2 = 0.0997$ $p = 0.6405$, $p = 0.6836$	6.46%	7.85%	11.57%	12.65%
Methyl (CH_3) $R^2 = 0.0653$, $R^2 = 0.1664$ $p = 0.7454$, $p = 0.5926$	0.89%	0.84%	0.25%	0.93%
Formate ($HCOO^-$)	10.57%	3.51%	24.18%	33.91%

R^2 = 0.0033, R^2 = 0.0124 p = 0.9428, p = 0.8889				
Methanol (CH_3OH) R^2 = 0.9418, R^2 = 0.9279 p = 0.0297, p = 0.0365	3.69%	0.47%	0.72%	1.31%
Acetate (CH_3COO^-) R^2 = 0.4217, R^2 = 0.3909 p = 0.3506, p = 0.3748	34.63%	15.97%	1.75%	2.10%

References

Bostick, K. W., Zimmerman, A. R., Goranov, A. I., Mitra, S., Hatcher, P. G., and Wozniak, A. S.: Biolability of fresh and photodegraded pyrogenic dissolved organic matter from laboratory-prepared chars, Journal of Geophysical Research: Biogeosciences, 126, 1-17, https://doi.org/10.1029/2020JG005981, 2021.~~Bostick, K. W., Zimmerman, A. R., Goranov, A. I., Mitra, S., Hatcher, P. G., and Wozniak, A. S.: Biolability of fresh and photodegraded pyrogenic dissolved organic matter from laboratory-prepared chars, ESSOAr [pre-print], https://doi.org/10.1002/essoar.10503766.1, 31 July 2020.~~

Chen, H. M., Stubbins, A., Perdue, E. M., Green, N. W., Helms, J. R., Mopper, K., and Hatcher, P. G.: Ultrahigh resolution mass spectrometric differentiation of dissolved organic matter isolated by coupled reverse osmosis-electrodialysis from various major oceanic water masses, Marine Chemistry, 164, 48-59, 10.1016/j.marchem.2014.06.002, 2014.

Helms, J. R., Stubbins, A., Ritchie, J. D., Minor, E. C., Kieber, D. J., and Mopper, K.: Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter, Limnology and Oceanography, 53, 955-969, https://doi.org/10.4319/lo.2008.53.3.0955, 2008.

Hertkorn, N., Benner, R., Frommberger, M., Schmitt-Kopplin, P., Witt, M., Kaiser, K., Kettrup, A., and Hedges, J. I.: Characterization of a major refractory component of marine dissolved organic matter, Geochimica et Cosmochimica Acta, 70, 2990-3010, https://doi.org/10.1016/j.gca.2006.03.021, 2006.

Koch, B. P., and Dittmar, T.: From mass to structure: An aromaticity index for high-resolution mass data of natural organic matter, Rapid Communications in Mass Spectrometry, 20, 926-932, https://doi.org/10.1002/rcm.2386, 2006.

Koch, B. P., and Dittmar, T.: From mass to structure: An aromaticity index for high-resolution mass data of natural organic matter (Erratum), Rapid Communications in Mass Spectrometry, 30, 1, https://doi.org/10.1002/rcm.7433, 2016.

Sleighter, R. L., Chen, H., Wozniak, A. S., Willoughby, A. S., Caricasole, P., and Hatcher, P. G.: Establishing a measure of reproducibility of ultrahigh-resolution mass spectra for complex mixtures of natural organic matter, Analytical Chemistry, 84, 9184-9191, https://doi.org/10.1021/ac3018026, 2012.

Sleighter, R. L., and Hatcher, P. G.: Molecular characterization of dissolved organic matter (DOM) along a river to ocean transect of the lower Chesapeake Bay by ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry, Marine Chemistry, 110, 140-152, 10.1016/j.marchem.2008.04.008, 2008.