

Article

# Untargeted Exometabolomics Provides a Powerful Approach to Investigate Biogeochemical Hotspots with Vegetation and Polygon Type in Arctic Tundra Soils

Mallory P. Ladd <sup>1,2</sup>, David T. Reeves <sup>1,2</sup>, Suresh Poudel <sup>2,3</sup>, Colleen M. Iversen <sup>4</sup>, Stan D. Wullschleger <sup>4</sup> and Robert L. Hettich <sup>2,\*</sup>

- <sup>1</sup> Bredeben Center for Interdisciplinary Research & Graduate Education, University of Tennessee, Knoxville, TN 37996, USA; mallory.ladd@gmail.com (M.P.L.); dreeves4@vols.utk.edu (D.T.R.)  
<sup>2</sup> Oak Ridge National Laboratory, Biosciences Division, Oak Ridge, TN 37830, USA; Suresh.Poudel@STJUDE.ORG  
<sup>3</sup> Genome Science & Technology, University of Tennessee, Knoxville, TN 37996, USA  
<sup>4</sup> Oak Ridge National Laboratory, Environmental Sciences Division, Climate Change Science Institute, Oak Ridge, TN 37830, USA; iversencm@ornl.gov (C.M.I.); wullschlegd@ornl.gov (S.D.W.)  
 \* Correspondence: hettichrl@ornl.gov; Tel.: +1-865-574-4968

## Abbreviation

6-MAP	4-amino-6-methyl-8-(2'-deoxy-β-D-ribofuranosyl)-7(8H)-pteridone
ACN	acetonitrile
ANOVA	analysis of variance
CID	collision-induced dissociation
DOM	dissolved organic matter
ESI	electrospray ionization
FA	formic acid
FC	fold change
FDR	false discovery rate
GUI	graphical user interface
HCP	high-centered polygon
HQFs	high quality features
IC	inorganic carbon
LC	liquid chromatography
LCP	low-centered polygon
LMW	low molecular weight
LOESS	locally estimated scatterplot smoothing
MeOH	methanol
MS	mass spectrometry
PCA	principal component analysis
PLS-DA	partial least-squares discriminant analysis
QC	quality control
RANSAC	random sample consensus
RT	retention time
S/N	signal-to-noise
TN	total nitrogen
TOC	total organic carbon
XIC	extracted ion chromatogram

**Citation:** Ladd, M.P.; Reeves, D.T.; Poudel, S.; Iversen, C.M.; Wullschleger, S.D.; Hettich, R.L. Untargeted exometabolomics provides a powerful approach to investigate biogeochemical hotspots with vegetation and polygon type in Arctic tundra soils. *Soil Syst.* **2021**, *5*, 10. <https://doi.org/10.3390/soilsystems5010010>

Academic Editor: Carlo Viti  
 Received: 29 December 2020  
 Accepted: 3 February 2021  
 Published: 9 February 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Materials and Methods

**Study site and sample description.** Soil cores ( $n = 4$ , organic horizon only, 10 cm diameter, ~15 cm depth) were obtained using a push-corer and a long knife from the Barrow Environmental Observatory, a continuous-permafrost, polygonal tundra landscape on the northern coastal plain of Alaska ( $71^{\circ} 16'N$ ,  $156^{\circ} 36'W$ ). Mineral soil was visually identified and removed by hand in the field along with any loose vegetative material. To examine the relationship between polygon, vegetation, and LWM DOM availability, two cores were collected from the center of a low-centered polygon (LCP) and two from the center of a high-centered polygon (HCP), where the aboveground vegetation was primarily either *Carex aquatilis* or *Eriophorum angustifolium*, two dominant plant species in these systems. Triplicate samples from each core were analyzed as biological replicates to ensure statistical power and enable comparative analyses. The cores were collected in late-August when the active layer had reached its maximal depth (~34 cm) [1]. There were no visible signs of cryoturbation in each of the cores. The mean air temperature for this region during August is  $4^{\circ}C$  and the mean annual precipitation is 10.74 cm [2]. Additional information about the study site and soil types has been described in detail previously [3]. Cores were immediately sealed in gallon freezer-bags (Ziplock), stored on blue ice for transport to a  $-20^{\circ}C$  freezer to slow microbial metabolic activity until field work was completed. Cores were then transported frozen from Alaska to Oak Ridge National Laboratory (ORNL) in Oak Ridge, Tennessee using blue ice and a sealed cooler and stored at  $-80^{\circ}C$  until processing.

Soil water content (Equation 1) measurements were made using a gravimetric soil moisture technique. Gravimetric analyses were completed by drying a subsample of soil (4 g) to constant weight in an oven at  $105^{\circ}C$  for 48 hours.

$$\% \text{ soil moisture} = \frac{\text{Fresh wet weight} - \text{dry weight}}{\text{dry weight}} * 100 \quad (1)$$

To obtain gross estimates of live root biomass (root weight, g), live roots (determined visually by color and roundness/diameter) were removed and set aside during homogenization. Homogenization was limited to 20 min to reduce human-derived variation in the number of roots removed from each soil. Roots were dried to constant weight at  $60^{\circ}C$  for 24 hours. Total organic carbon (TOC) and total nitrogen (TN) data were collected in triplicate on a Shimadzu TOC-L CSH/CSN analyzer (Columbia, MD). Briefly, a subsample of soil (2 g) or soil pore water (24 mL), is introduced to the instrument where it then transferred to a combustion tube. For TOC analyses, both pure and an acidified sample are analyzed to obtain a total carbon (TC) and an inorganic carbon (IC) measurement, respectively, which can then be used to calculate TOC ( $TOC = TC - IC$ ). A carrier gas (zero-carbon air) flows at 150 mL/min to the combustion tube, which has been filled with an oxidation catalyst (platinum) and is heated to  $680^{\circ}C$ . The TC or IC of a sample is combusted into  $CO_2$  which is then carried to a dehumidifier, where it is cooled, dehydrated, and detected using nondispersive infrared gas analysis (NDIR). The analog detection signal of the NDIR forms a peak which is proportional to the TC concentration of the sample. Using a standard TC solution, a calibration curve is generated, and unknown TC concentrations may be calculated. For TN analysis, samples are introduced into the combustion tube packed with a catalyst (platinum) and the furnace temperature is set to  $720^{\circ}C$ , creating nitrogen monoxide (NO) gas. Zero-carbon air is used to carry NO to the chemiluminescence analyzer where the NO reacts with ozone ( $O_3$ ) creating products that are then measured photo-electrically generating a peak proportional to the total nitrogen concentration in the sample. Unknown concentrations are determined using a calibration curve as well.

**Chemicals.** Mobile phase solvents included degassed LC/MS-grade water ( $H_2O$ ), acetonitrile (ACN), methanol (MeOH), and isopropyl alcohol (IPA) which were obtained from EMD Millipore (Billerica, MA, USA). Mobile phase additives included ammonium

acetate (NH<sub>4</sub>Ac), ammonium hydroxide (NH<sub>4</sub>OH), and formic acid (FA) which were obtained from Sigma-Aldrich (St. Louis, MO, USA).

**Instrumentation and LC/MS data collection.** Samples and controls were thawed and prepared immediately prior to injection by adding either FA or NH<sub>4</sub>OH (0.1 %) to help with ionization in positive- or negative-ion mode, respectively. Each sample was manually injected directly onto the columns using a 300 nL fused-silica loop, and nano-flow rates were achieved using a split-flow setup prior to the injection loop. To control for instrument drift, the mass spectrometer was externally calibrated every two days or before switching columns or polarities using a mixture of the peptide MRFA, caffeine, and Ultramark 1621 in MeOH, ACN, and acetic acid for positive-ion mode or a mixture of sodium taurocholate, Ultramark 1621, and sodium dodecyl sulfate in the same solvents for negative-ion mode (Pierce, ThermoFisher Scientific). The ESI source voltage and temperature were optimized to 2.2 kV or 2.8 kV and 225 °C or 275 °C for positive- or negative-ion mode, respectively. Fragmentation spectra were acquired in a “top ten” data-dependent mode where the ten most abundant precursor ions in each full scan were isolated (2 *m/z* isolation width, 15,000 resolving power) for fragmentation by collision-induced dissociation (CID) with He<sub>(g)</sub> at a normalized CID energy of 30 (unitless). Dynamic exclusion of MS<sup>1</sup> ions selected for fragmentation was applied for 60 seconds to improve detection of low abundant or coeluting features. A charge stage rejection for multiply charged precursors was employed and two microscans were averaged for every full MS<sup>1</sup> and MS<sup>2</sup> spectrum to reduce spectral complexity and improve reproducibility of acquired spectra.

**Untargeted LC/MS data processing.** Raw LC/MS data were subjected to peak picking, alignment, and normalization using MZmine2 (v2.30) [4,5]. This software is open-source and has a user-friendly graphical user interface (GUI) with separate modules for each data processing step, but also includes a batch-processing mode, maximizing the accessibility of the software’s capabilities to new users or experienced analysts alike. Prior to statistical analyses, it is important in untargeted analyses to be able to detect as many small, but real analyte signals as possible. Here, differentiating between true and false signals was accomplished by first optimizing three parameters in the MZmine peak extraction algorithm—minimum peak height, MS<sup>1</sup> tolerance, and RT window. These parameters are optimized by manually inspecting the accuracy of the peak assignment for a subset of the data.

Precursor ions (MS<sup>1</sup>) that were selected for fragmentation (MS<sup>2</sup>) and had an intensity above a specified noise level (S/N > 3) were identified with the MS/MS peak list builder and user-defined parameters (e.g. +/- 0.005 *m/z* or 10 ppm MS<sup>1</sup> window). Next, chromatograms were built using the peak extender module which searches for the same peak (MS<sup>1</sup> and MS<sup>2</sup>) in both directions of the retention time (RT) apex within a given *m/z* and RT tolerance (e.g. +/- 10 ppm, +/- 2 min), resulting in an assigned peak area. Isotopic peaks (*i.e.* <sup>13</sup>C natural abundance ion, mass difference of a neutron = 1.0033 Da) were then removed with the isotopic peaks grouper module using a *m/z* and RT tolerance in order to avoid errors with relative quantitation and annotation. A slightly wider than normal isotope RT tolerance was used here in order to adequately account for potential peak width widening specifically within HILIC runs. During the ESI process, while less likely than with other ionization techniques, in-source fragmentation can occur, along with the formation of non-proton adducts with Na<sup>+</sup>, K<sup>+</sup>, or NH<sub>4</sub><sup>+</sup> for example, or complexes that co-elute with analytes of interest. Here, fragments were identified in MZmine by comparing peak lists with MS<sup>2</sup> scan data (e.g. same *m/z* within +/- 5 ppm and same RT +/- 0.1 min), while adducts were identified in MZmine by the mass difference between the original ion and the adduct being equal to the mass selected by the user (+/- 5 ppm from 22.9892 *m/z* for a Na<sup>+</sup> adduct) and having a matching RT (+/- 0.1 min). Finally, complexes were identified in MZmine by searching for peaks with the same RT time (+/- 0.1 min) that add together to make the ion complex *m/z* (+/- 5 ppm). To help reduce any chromatogram shifts

that would impact annotation, but include features whose RTs had shifted slightly between extraction replicates, peaks from the same chromatographic phase and ionization mode were aligned ( $\pm 5$  ppm,  $\pm 2$  min RT) based on 10 iterations and at least a 25 % match score using the nonlinear, random sample consensus (RANSAC) algorithm.[5,6] Aligned peak lists were exported to .csv files for data filtering procedures. It's important to note that the adducts, complex, and fragments identified in the identification module of MZmine can be removed from the dataset at any point in data filtering process. For the sake of evaluating the technique, here, they were not removed in order to evaluate the proportion of which may be annotated as LMW DOM metabolites by database searching as well.

**Data filtering, normalization, and statistical analyses.** Multiple conservative LC/MS-based metabolomic data processing techniques were applied here. Integrated LC peak areas were obtained from the aligned extracted ion chromatograms (XICs), normalized to per gram dry soil to account for moisture variations between samples, and then  $\log_2$ -transformed for ease of data interpretation. To control for systematic variation between samples and remove intragroup batch effects, the  $\log_2$ -transformed peak areas were also normalized to pooled-sample QCs using a common approach—QC-RLSC (robust LOESS signal correction) [7] with two scaling factor techniques, LOESS (locally estimated scatterplot smoothing) and median-centering, all completed in the freely-available InfernoRDN and R environments [8]. By including controls and daily technical blanks, artifact signals that originated from sample collection, preparation, or analysis that were above a specified noise level could then be readily identified and manually removed, decreasing the false discovery rate (FDR) of the technique [9,10].

While there are many different methods for normalizing metabolomics data, each comes with various drawbacks and tradeoffs (*i.e.* bias-variance trade-off) and no single approach perfectly describes all the unwanted variation associated with an experiment, which is why it is important to consider the experimental design and aims of the study when optimizing a normalization approach [11]. For example, while normalizing to an internal standard that is specific to each compound-of-interest (targeted analyses) or to a mixed internal standard with compounds from multiple classes for untargeted analyses are alternative normalization approaches commonly used in metabolomics analyses, these require the introduction of several external compounds to the sample, which not only further complicate the chromatogram and mass spectrum, but could also alter the composition of the sample via chemical reactions. Thus, here we chose to use a pooled quality control consisting of equal volumes of all 36 samples and a single internal standard, 4-amino-6-methyl-8-(2'-deoxy- $\beta$ -D-ribofuranosyl)-7(8H)-pteridone (6-MAP), that ionized well in both positive and negative ion mode.

After filtering the data set to obtain only the high-quality features (HQFs) and their corresponding peak area data, missing values were imputed for statistical analyses by randomly selecting numbers from a normal distribution near the limit of detection (width = 0.3, downshift = 1.8-2.3) using the freely-available Perseus software [12]. Variation between replicate extracts to assess reproducibility were analyzed using Pearson's correlations performed with JMP Pro (v13.1 SAS Institute). To analyze the variation due to polygon type or vegetation, principal component analyses (PCA) were first used to visualize the overall variation across the untargeted datasets [13]. PCA is an unsupervised, data dimension-reduction technique that plots the weighted-sum of the contribution of a set of LMW DOM features within a sample to a principal component and compares that to all the other samples. While PCA can be used as a multivariate statistical analysis, it suffers from the multi-collinearity problem that is common with metabolomics datasets, in that they generally have more dependent variables (*i.e.* metabolites, in the hundreds or thousands) than independent variables (*i.e.* biological conditions, in the tens). An alternative statistical approach that is frequently applied in metabolomic datasets is that of partial

least squares discriminant analysis (PLS-DA) which alleviates the independent-to-dependent ratio issue [14]. However, PLS-DA is a *supervised* technique, in that it plots the variation in the dataset after first considering the correlation between the dependent and independent variables. Thus, PCA was used here to first visualize the overall variation across the untargeted datasets.

Then, to determine differentially-abundant LMW DOM features, Student's t-test was used to compare profiles between cores of the same polygon type or vegetation and analysis of variance (ANOVA) was performed using the Python SciPy library [15]. Tukey's range test was used as a post-hoc analysis to compare all possible pairs and identify abundance differences greater than the expected standard error between the HCP and LCP cores or the *Carex* and *Eriophorum* cores. Any feature with a  $\log_2$  fold change  $> 2$  and a p-value  $< 0.05$  was considered significant, but we also explored tighter parameters ( $\log_2$  fold change  $> 4$ , p-value  $< 0.001$ ) to highlight LMW DOM features that were highly significant. Because pairwise comparisons by t-test can lead to a multiple-testing error with metabolomic datasets, *volcano plots*—which consider the fold change (FC) between two conditions—were used to identify significant features that passed both a p-value threshold and a FC threshold. Finally, two-way hierarchically-clustered heat maps using the Ward agglomerative technique were used to visualize these variations and select clusters of features that varied similarly across the dataset for annotation. Volcano plots and heatmaps were generated in Perseus, and PCAs were produced in the InfernoRDN environment.

**Feature annotation.** Annotation of features that were consistently observed and significantly differentially-abundant due to polygon type or vegetation was carried out in a three-step procedure. First, features ([M+H]<sup>+</sup> or [M-H]<sup>-</sup> ions) were searched against multiple freely-available online databases using high mass accuracy measurements (precursor mass tolerance of 5 ppm) within MZmine and using the MetaboSearch tool [16]. Databases included KEGG [17], METLIN [18], MMCD [19], PubChem [20], HMDB [21], LipidMaps [22], or Plant Cyc [23]. While it depends on the database size, this first filter is the most powerful and generally can remove up to 99.9 % of false candidates [24]. Second, putative chemical formulas were assigned using the MZmine elemental formula assignment module and the following criteria established using Kind and Fiehn's "Seven Golden Rules" and parameters modified from Kujawinski and Behn's compound identification algorithm (CIA) for small molecules [25-27]: mass measurement error of  $< 5$  ppm, taking into account the presence of C<sub>1-100</sub>, H<sub>3-100</sub>, N<sub>0-30</sub>, O<sub>1-50</sub>, P<sub>0-3</sub>, S<sub>0-3</sub>, double bond-equivalents (Equation 2), aromaticity index (AI, Equation 3), and elemental ratio heuristics including  $0.1 \leq H/C \leq 6$ ,  $N/C \leq 4$ ,  $O/C \leq 3$ ,  $P/C \leq 2$ , and  $S/C \leq 3$ .

$$DBE = 1 + C - 0.5H + 0.5N + 0.5P \quad (2)$$

$$AI = \frac{1 + C - 0.5O - S - 0.5H}{C - 0.5O - S - N - P} \quad (3)$$

Van Krevelen plots were used to visualize the formula assignments. Boxes overlaid on the plots indicate assigned biochemical classes [28,29]: lipids ( $O/C < 0.3$ ,  $H/C > 1.7$ ), peptides, amino acids, and amino sugars ( $0.3 < O/C < 0.7$ ,  $H/C > 1.5$ ), carbohydrates ( $O/C > 0.7$ ,  $H/C > 1.5$ ), unsaturated hydrocarbons ( $O/C < 0.1$ ,  $0.7 < H/C < 1.7$ ), lignins ( $0.1 < O/C < 0.7$ ,  $0.7 < H/C < 1.7$ ), tannins ( $O/C > 0.7$ ,  $H/C < 1.5$ ), and phenolics/condensed aromatics ( $O/C < 0.7$ ,  $H/C < 0.7$ ).

When multiple candidate formulas were returned, to ensure that an objective choice was made, we consistently chose the formula with the lowest error, lowest number of heteroatoms, and if there was a phosphorus present, at least three oxygen atoms must have also been present in the formula [30]. Third, compounds that matched to multiple hits in a database were manually scrutinized in an iterative approach by assessing high-resolution mass spectral data for consistent fragmentation profiles, or by using the similarity matching tool in MZmine, to filter out false candidates and annotate unknown (unmatched) features.

---

It is important to note here that while we included an *annotation* step, it was outside the scope of this study to *identify* the LMW DOM features by matching to authentic standards as that would limit our analytical window to only metabolites that have been synthesized. Due to the complexity of this analyte pool, most of the features detected are likely “unknowns”, and authentic standards are frequently unavailable. For the aim of distinguishing a profile of features (known or unknown) that were differentially-abundant across space, with the ultimate goal of linking that chemical profile to biological processes (*i.e.* methanogenesis) or having it act as an indicator of C vulnerability, high-mass accuracy MS<sup>1</sup> and MS<sup>2</sup> annotations and putative identifications by database matching or elemental formula assignment are sufficient.

**Table S1.** Polygon soil core sample summary and corresponding TOC, TN, TC, C:N, and dry root weight results.

Extract Number	Site	Polygon Type	Vegetation	Water Content (%)	g H <sub>2</sub> O/g dry soil	TOC (%)	TN (%)	TC (%)	OC:N	Dry Root Wt (g)
1	A	LCP	Carex							
2	A	LCP	Carex	82.8	4.84	41.415	2.570	48.290	16.117	0.1444
3	A	LCP	Carex							
4	A	LCP	Carex							
5	A	LCP	Carex	80.3	4.07	42.624	2.228	46.966	19.131	0.1814
6	A	LCP	Carex							
7	A	LCP	Carex							
8	A	LCP	Carex	79.6	3.91	42.104	2.241	46.016	18.792	0.0778
9	A	LCP	Carex							
10	B	HCP	Carex							
11	B	HCP	Carex	73.4	2.75	41.521	2.479	46.621	16.750	0.3746
12	B	HCP	Carex							
13	B	HCP	Carex							
14	B	HCP	Carex	72.1	2.58	43.464	2.567	45.209	16.930	0.0605
15	B	HCP	Carex							
16	B	HCP	Carex							
17	B	HCP	Carex	73.4	2.76	37.334	2.1915	41.741	17.0358	0.1694
18	B	HCP	Carex							
19	A	LCP	Eriophorum							
20	A	LCP	Eriophorum	85.1	5.69	34.253	1.931	47.441	17.742	1.0316
21	A	LCP	Eriophorum							
22	A	LCP	Eriophorum							
23	A	LCP	Eriophorum	83.8	5.16	35.809	2.242	47.098	15.971	0.5866
24	A	LCP	Eriophorum							

25	A	LCP	Eriophorum							
26	A	LCP	Eriophorum	76.5	3.26	38.673	2.308	43.615	16.755	0.1730
27	A	LCP	Eriophorum							
28	B	HCP	Eriophorum	75.6	3.09	39.803	2.189	47.619	18.185	1.1620
29	B	HCP	Eriophorum							
30	B	HCP	Eriophorum	75.6	3.09	39.803	2.189	47.619	18.185	1.1620
31	B	HCP	Eriophorum	73.8	2.82	39.554	2.357	46.203	16.782	0.3173
32	B	HCP	Eriophorum							
33	B	HCP	Eriophorum	73.8	2.82	39.554	2.357	46.203	16.782	0.3173
34	B	HCP	Eriophorum	70.4	2.38	41.026	2.431	44.054	16.873	0.0446
35	B	HCP	Eriophorum							
36	B	HCP	Eriophorum	70.4	2.38	41.026	2.431	44.054	16.873	0.0446

**Table S2.** Optimized mobile phase conditions and additives for each LC phase and MS polarity, injection volume, and flow rates.

	Solvent A			Solvent B		
<b>HILIC (+)</b>	<b>60 % ACN, 40 % 5 mM NH<sub>4</sub>Ac, 0.1% FA</b>			<b>95 % ACN, 5 % 5 mM NH<sub>4</sub>Ac, 0.1 % FA</b>		
<b>HILIC (-)</b>	100 % 5 mM NH <sub>4</sub> Ac, 0.1 % NH <sub>4</sub> OH			95 % ACN, 5 % 5 mM NH <sub>4</sub> Ac, 0.1 % NH <sub>4</sub> OH		
<b>RP (+)</b>	95 % H <sub>2</sub> O, 5 % ACN, 0.1 % FA			70 % ACN, 30 % H <sub>2</sub> O, 0.1 % FA		
<b>RP (-)</b>	90 % H <sub>2</sub> O, 10 % IPA, 1 mM NH <sub>4</sub> OH			80 % ACN, 10 % H <sub>2</sub> O, 10 % IPA, 1 mM NH <sub>4</sub> OH		
<b>Injection Volume</b>	300 nL	<b>Flow Rate at Pump</b>	0.100 mL/min	<b>Flow Rate at Tip</b>	~250 nL/min	

**Table S3.** Optimized gradient conditions for nano-LC separations, for positive- and negative-MS modes on C18-RP and ZIC-pHILIC columns.

C18 Reversed-Phase				ZIC-pHILIC			
Positive		Negative		Positive		Negative	
time, min	% B	time, min	% B	time, min	% A	time, min	% A
0.0	2	0.0	25	0.0	0	0.0	0
3.0	2	3.0	25	3.0	0	3.0	0
23.0	100	23.0	100	23.0	100	23.0	30
28.0	100	28.0	100	28.0	100	28.0	30
33.0	2	33.0	25	30.0	80	30.0	60
40.0	2	40.0	25	35.0	80	35.0	60
				40.0	0	40.0	0
				45.0	0	45.0	0

**Table S4.** MZmine parameters used for the peak detection modules applied in the analysis of the polygonal tundra soil organic horizons.

<b>Peak Detection Methods.</b>				
<b>Mass Detection</b>	<b>HILIC (+)</b>	<b>HILIC (-)</b>	<b>RP (+)</b>	<b>RP (-)</b>
RT window:	Auto range	Auto range	Auto range	Auto range
MS level:	1 and 2	1 and 2	1 and 2	1 and 2
Polarity:	+	-	+	-
Spectrum type:	centroided	centroided	centroided	centroided
MS1 noise level:	1.00E+04	2.00E+05	5.00E+03	1.00E+03
MS2 noise level:	5.00E+02	4.00E+02	5.00E+02	1.00E+02
<b>MS/MS Peak List Builder</b>				
RT window:	Auto range	Auto range	Auto range	Auto range
MS level:	2	2	2	2
Polarity:	+	-	+	-
Spectrum type:	centroided	centroided	centroided	centroided
m/z window	0.01	0.01	0.01	0.01
Time window	61 min	56 min	41 min	41 min
<b>Peak Extender</b>				
m/z tolerance:	0.005 m/z or 10 ppm			
Min height	1.00E+04	1.00E+02	1.00E+03	1.00E+03

**Table S5.** MZmine parameters used for the peak list generation modules applied in the analysis of the polygonal tundra soil organic horizons.

<b>Peak List Generation Methods</b>				
<b>Isotopic Peaks Grouper</b>	<b>HILIC (+)</b>	<b>HILIC (-)</b>	<b>RP (+)</b>	<b>RP (-)</b>
m/z tolerance:	0.005 m/z or 10 ppm			
RT tolerance:	1.0 min	1.0 min	1.0 min	1.0 min
Monotonic shape:	Y	Y	Y	Y
Maximum charge:	1	1	1	1
Representative isotope:	Most intense	Most intense	Most intense	Most intense
<b>Duplicate Peaks Filter</b>				
m/z tolerance:	0.001 m/z or 5 ppm			
RT tolerance:	0.25 min	0.25 min	0.25 min	0.25 min
<b>RANSAC Aligner</b>				
m/z tolerance:	0.005 m/z or 10 ppm			
RT tolerance:	61 min	56 min	41 min	41 min
RT tolerance after correction:	20 min	30 min	20 min	20 min
RANSAC Iterations:	0 (model optimized)	0 (model optimized)	0 (model optimized)	0 (model optimized)
Minimum number of points:	25%	25%	30%	30%
<b>Gap Filling</b>				
m/z tolerance:	0.005 m/z or 10 ppm			
RT tolerance:	61 min	56 min	41 min	41 min
<b>Annotation</b>				
m/z tolerance:	0.005 m/z or 10 ppm			
m/z vs RT balance:	0.2 min	0.2 min	0.2 min	0.2 min

---

Max fragment peak height:	80%	80%	80%	80%
Min MS2 peak height:	5.00E+02	1.00E+02	5.00E+02	1.00E+02
Max complex peak height:	50%	50%	50%	50%
Max adduct peak height:	50%	50%	50%	50%
Online databases searched:	KEGG, PubChem, HMDB, LipidMaps, PlantCyc			

---

**Table 6.** “Cluster 1”, listing the differentially-abundant LMW DOM features found in lower relative abundance in the *Eriophorum* – HCP core.

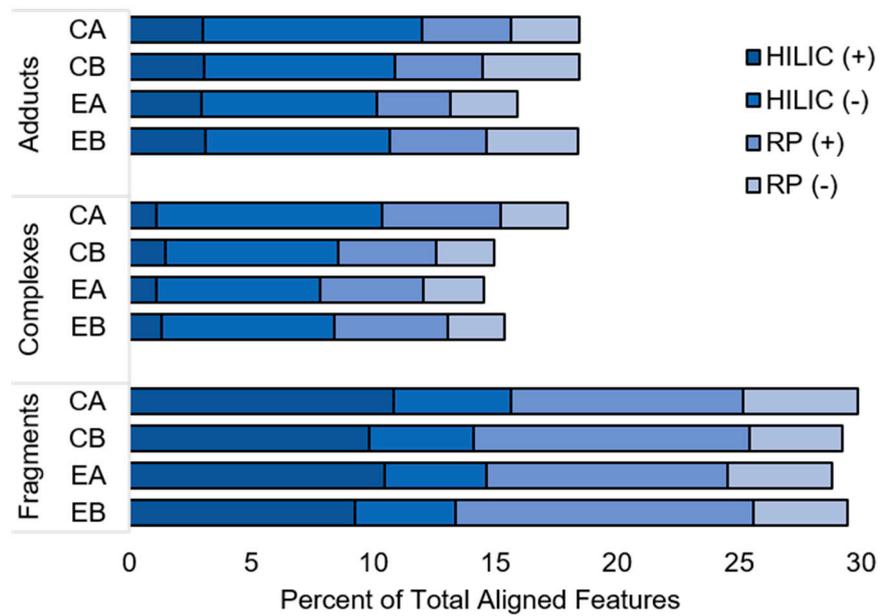
<i>m/z</i>	Predicted Formula	Mass error (ppm)	Compound Class	Database Annotation	Database Formula	Database Compound Class	MW
321.0933	C <sub>11</sub> H <sub>18</sub> N <sub>2</sub> O <sub>9</sub>	-2	carbohydrate	-	-	-	-
275.9782	C <sub>11</sub> H <sub>3</sub> NO <sub>8</sub>	-1.3	tannin	-	-	-	-
247.9740	C <sub>11</sub> H <sub>6</sub> N <sub>4</sub> O <sub>19</sub>	0	tannin	-	-	-	-
325.1183	C <sub>12</sub> H <sub>19</sub> N <sub>6</sub> O <sub>3</sub> P	-0.1	lignin	-	-	-	-
191.5355	C <sub>12</sub> H <sub>19</sub> NO <sub>13</sub>	-0.1	carbohydrate	-	-	-	-
380.0831	C <sub>13</sub> H <sub>19</sub> NO <sub>12</sub>	-0.6	tannin	-	-	-	-
265.0606	C <sub>15</sub> H <sub>10</sub> N <sub>2</sub> O <sub>3</sub>	-3.5	aromatic	6-acetophenazine-1-carboxylic acid	C <sub>15</sub> H <sub>10</sub> N <sub>2</sub> O <sub>3</sub>	aromatic	266.069
281.0920	C <sub>16</sub> H <sub>14</sub> N <sub>2</sub> O <sub>3</sub>	-3	lignin	-	-	-	-
311.1029	C <sub>17</sub> H <sub>16</sub> N <sub>2</sub> O <sub>4</sub>	-1.6	lignin	-	-	-	-
110.9824	C <sub>17</sub> H <sub>36</sub> N <sub>2</sub> O <sub>3</sub>	0.7	lipid	-	-	-	-
511.4389	C <sub>17</sub> H <sub>50</sub> N <sub>16</sub> O <sub>2</sub>	2.6	lipid, aliphatic	-	-	-	-
337.0826	C <sub>18</sub> H <sub>14</sub> N <sub>2</sub> O <sub>5</sub>	0.1	lignin	-	-	-	-
145.0889	C <sub>18</sub> H <sub>27</sub> N <sub>15</sub> O <sub>6</sub>	-0.3	lignin	-	-	-	-
225.9641	C <sub>19</sub> H <sub>2</sub> O <sub>14</sub>	-3.6	tannin	-	-	-	-
226.9553	C <sub>2</sub> H <sub>4</sub> N <sub>4</sub> O <sub>5</sub> S <sub>2</sub>	1.3	carbohydrate	-	-	-	-
272.9246	C <sub>2</sub> H <sub>4</sub> N <sub>4</sub> O <sub>6</sub> P <sub>2</sub> S	-2.8	carbohydrate	-	-	-	-
214.8694	C <sub>2</sub> HO <sub>4</sub> PS <sub>3</sub>	-3.5	tannin	-	-	-	-
267.0913	C <sub>3</sub> H <sub>12</sub> N <sub>10</sub> O <sub>5</sub>	-0.8	carbohydrate, aliphatic	-	-	-	-
253.1114	C <sub>3</sub> H <sub>14</sub> N <sub>10</sub> O <sub>4</sub>	-1.8	carbohydrate, aliphatic	-	-	-	-
223.9617	C <sub>3</sub> H <sub>3</sub> N <sub>3</sub> O <sub>7</sub> S	-0.8	tannin	-	-	-	-
246.9834	C <sub>3</sub> H <sub>4</sub> N <sub>8</sub> O <sub>2</sub> S <sub>2</sub>	3.4	lignin	-	-	-	-
243.0215	C <sub>3</sub> H <sub>8</sub> N <sub>4</sub> O <sub>9</sub>	0.3	carbohydrate, aliphatic	-	-	-	-
268.9273	C <sub>4</sub> H <sub>3</sub> N <sub>2</sub> O <sub>8</sub> PS	-0.6	tannin	-	-	-	-
231.9430	C <sub>4</sub> H <sub>3</sub> N <sub>5</sub> OS <sub>3</sub>	1.4	lignin	-	-	-	-
145.0621	C <sub>5</sub> H <sub>10</sub> N <sub>2</sub> O <sub>3</sub>	-3.7	protein	alanine-glycine	C <sub>5</sub> H <sub>10</sub> N <sub>2</sub> O <sub>3</sub>	protein	146.069
117.0561	C <sub>5</sub> H <sub>10</sub> O <sub>3</sub>	3.4	protein	2-hydroxy-3-methylbutyric acid	C <sub>5</sub> H <sub>10</sub> O <sub>3</sub>	metabolite	118.063
240.0767	C <sub>5</sub> H <sub>15</sub> N <sub>5</sub> O <sub>4</sub> S	-3.6	carbohydrate, aliphatic	(3Z)-3-(1H-imidazol-5-ylmethylene)-5-methoxy-1H-indol-2(3H)-one	C <sub>13</sub> H <sub>11</sub> N <sub>3</sub> O <sub>2</sub> *	aromatic, protein	241.085

206.9522	C <sub>5</sub> H <sub>5</sub> O <sub>5</sub> PS	-0.1	tannin	-	-	-	-
279.0931	C <sub>5</sub> H <sub>8</sub> N <sub>14</sub> O	2.3	lignin	-	-	-	-
206.0707	C <sub>6</sub> H <sub>14</sub> N <sub>3</sub> O <sub>3</sub> P	3.5	protein	-	-	-	-
429.1308	C <sub>6</sub> H <sub>18</sub> N <sub>14</sub> O <sub>9</sub>	0.7	carbohydrate, aliphatic	-	-	-	-
415.1512	C <sub>6</sub> H <sub>20</sub> N <sub>14</sub> O <sub>8</sub>	0.4	carbohydrate, aliphatic	-	-	-	-
231.9466	C <sub>6</sub> H <sub>4</sub> NO <sub>5</sub> PS	-3.8	tannin	-	-	-	-
211.0028	C <sub>6</sub> H <sub>5</sub> N <sub>4</sub> O <sub>3</sub> P	0.8	lignin	-	-	-	-
204.9729	C <sub>6</sub> H <sub>7</sub> O <sub>4</sub> PS	-0.3	lignin	-	-	-	-
350.8737	C <sub>6</sub> HN <sub>4</sub> O <sub>6</sub> PS <sub>3</sub>	4	tannin	-	-	-	-
191.0535	C <sub>7</sub> H <sub>13</sub> NO <sub>3</sub> S	-4.3	protein	-	-	-	-
253.0968	C <sub>7</sub> H <sub>18</sub> N <sub>4</sub> O <sub>4</sub> S	-2.7	protein, aliphatic	-	-	-	-
400.8702	C <sub>7</sub> H <sub>2</sub> N <sub>2</sub> O <sub>12</sub> S <sub>3</sub>	1.3	tannin	-	-	-	-
400.8711	C <sub>7</sub> H <sub>2</sub> N <sub>2</sub> O <sub>12</sub> S <sub>3</sub>	3.5	tannin	-	-	-	-
192.0527	C <sub>7</sub> H <sub>7</sub> N <sub>5</sub> O <sub>2</sub>	0.1	lignin	glucuronamide	C <sub>6</sub> H <sub>11</sub> NO <sub>6</sub> *	carbohydrate	193.059
347.8884	C <sub>7</sub> HN <sub>3</sub> O <sub>8</sub> P <sub>2</sub> S	-0.7	tannin	-	-	-	-
367.8842	C <sub>8</sub> H <sub>3</sub> NO <sub>10</sub> S <sub>3</sub>	-1.1	tannin	-	-	-	-
416.8452	C <sub>8</sub> H <sub>3</sub> O <sub>12</sub> PS <sub>3</sub>	0.2	tannin	-	-	-	-
416.8453	C <sub>8</sub> H <sub>3</sub> O <sub>12</sub> PS <sub>3</sub>	0.4	tannin	-	-	-	-
180.0653	C <sub>9</sub> H <sub>11</sub> NO <sub>3</sub>	-3.8	lignin	-	-	-	-
440.8636	C <sub>9</sub> H <sub>2</sub> N <sub>2</sub> O <sub>13</sub> S <sub>3</sub>	-2.3	tannin	-	-	-	-
76.0592	C <sub>9</sub> H <sub>9</sub> N <sub>2</sub> O <sub>2</sub>	0.6	lignin	4-ethoxy carbonyl benzenediazonium	C <sub>9</sub> H <sub>9</sub> N <sub>2</sub> O <sub>2</sub>	aromatic	177.066
89.0358	no hit	-	-	N-(hydroxy methyl)urea	C <sub>2</sub> H <sub>6</sub> N <sub>2</sub> O <sub>2</sub>	metabolite	90.0429
128.0724	no hit	-	-	6-carboxypiperidine	C <sub>6</sub> H <sub>11</sub> NO <sub>2</sub>	protein	129.079
138.0572	no hit	-	-	3-amino-2,3-dihydro benzoic acid	C <sub>7</sub> H <sub>9</sub> NO <sub>2</sub>	protein	139.063
218.1063	no hit	-	-	(2Z)-2-methyl-4-(9H-purine-6-ylamino)-2-buten-1-ol	C <sub>10</sub> H <sub>13</sub> N <sub>5</sub> O	plant hormone	219.112
94.9664	no hit	-	-	-	-	-	-
94.9666	no hit	-	-	-	-	-	-
102.0569	no hit	-	-	-	-	-	-
103.0540	no hit	-	-	-	-	-	-
110.9594	no hit	-	-	-	-	-	-
110.9765	no hit	-	-	-	-	-	-
112.0741	no hit	-	-	-	-	-	-

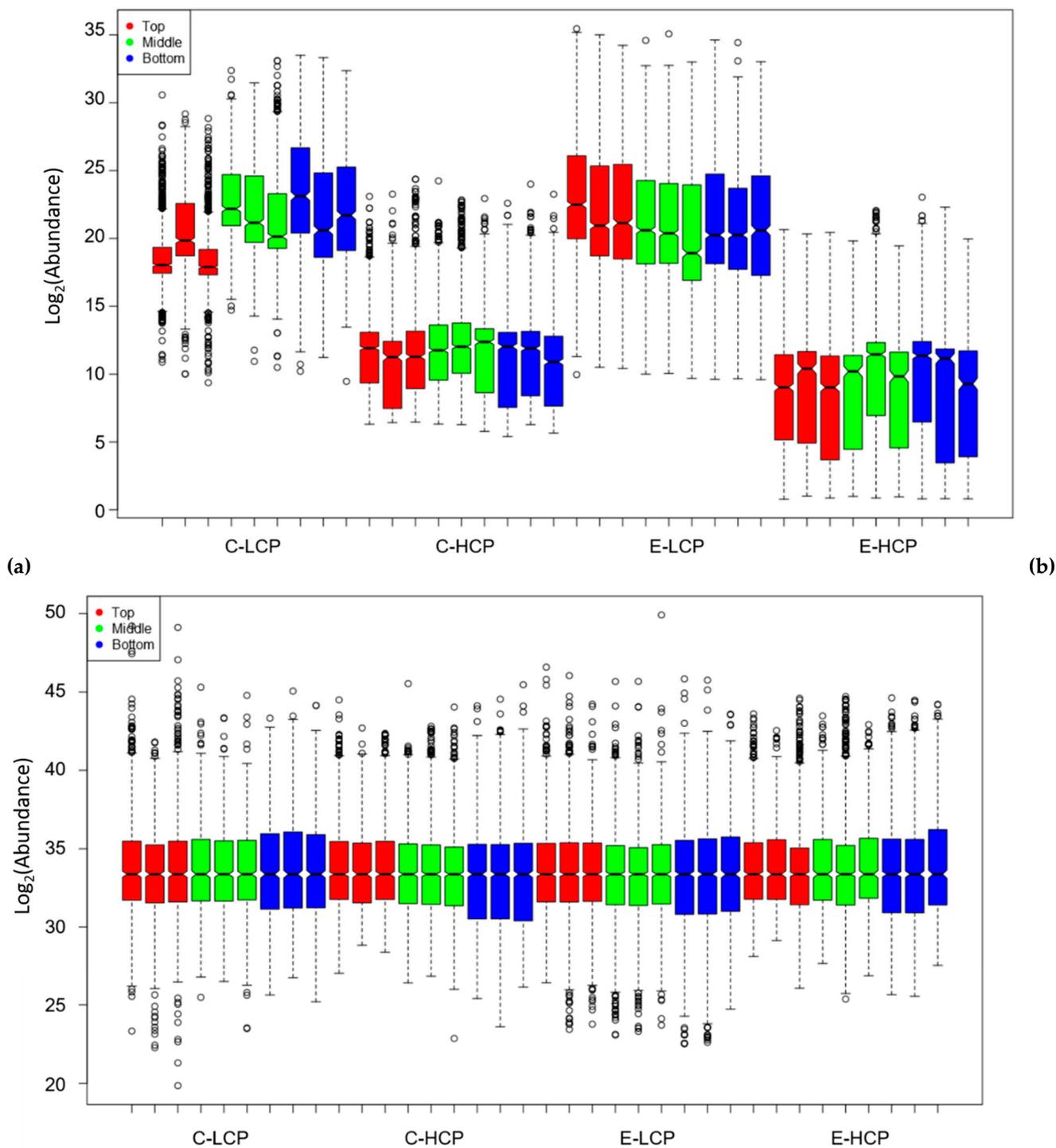
---

119.9483	no hit	-	-	-	-	-	-	-
124.9858	no hit	-	-	-	-	-	-	-
127.0539	no hit	-	-	-	-	-	-	-
133.0065	no hit	-	-	-	-	-	-	-
134.9178	no hit	-	-	-	-	-	-	-
147.0638	no hit	-	-	-	-	-	-	-
148.0022	no hit	-	-	-	-	-	-	-
149.9970	no hit	-	-	-	-	-	-	-
176.0906	no hit	-	-	-	-	-	-	-
196.9030	no hit	-	-	-	-	-	-	-
216.9125	no hit	-	-	-	-	-	-	-
236.8647	no hit	-	-	-	-	-	-	-
294.8031	no hit	-	-	-	-	-	-	-
324.7715	no hit	-	-	-	-	-	-	-
416.7706	no hit	-	-	-	-	-	-	-
488.8209	no hit	-	-	-	-	-	-	-

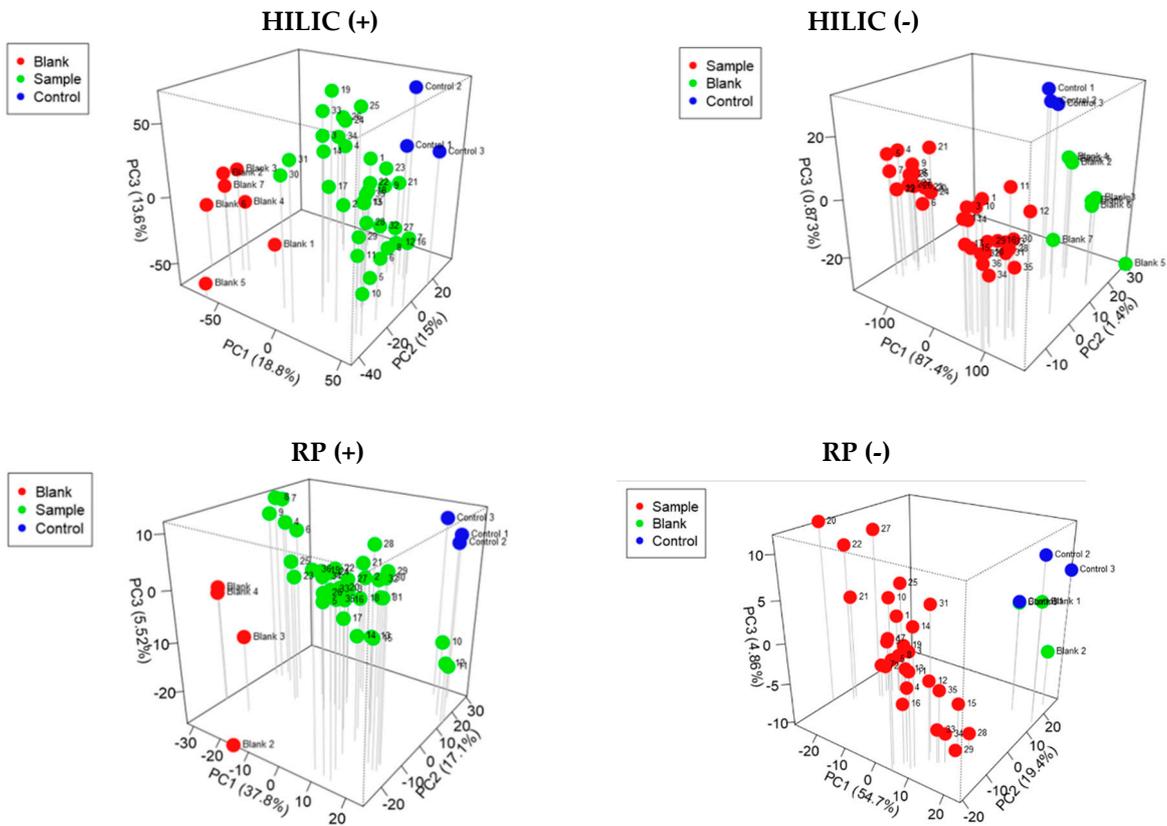
---



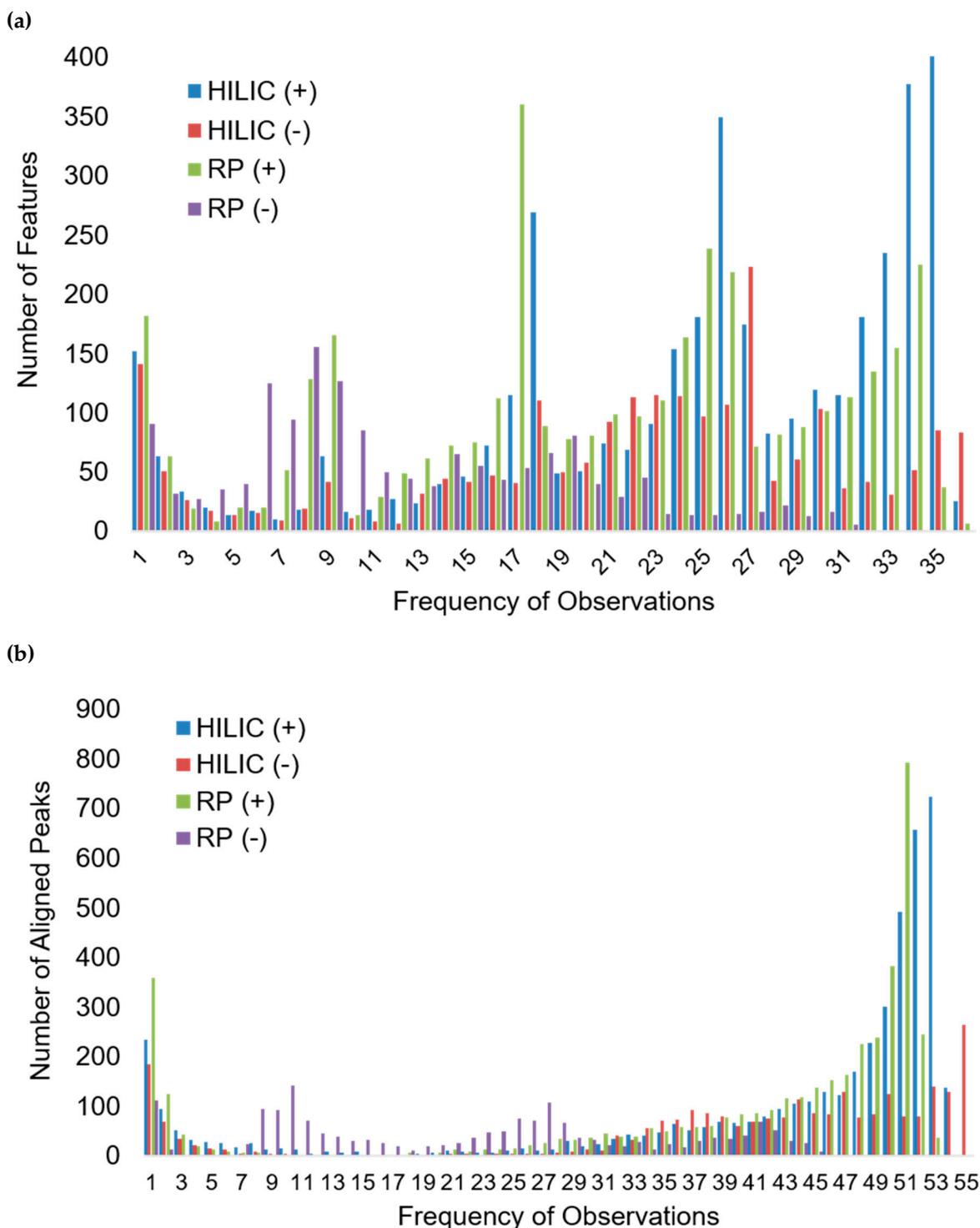
**Figure S1:** Percent aligned peaks that were annotated by MZmine as an adduct, complex, or fragment of another peak within 10 ppm and  $\pm 0.1$  min RT for each core grouped by LC/MS condition. C = *Carex*, E = *Eriophorum*, A = Site A or Low-Centered Polygon (LCP), and B = Site B or High-Centered Polygon (HCP).



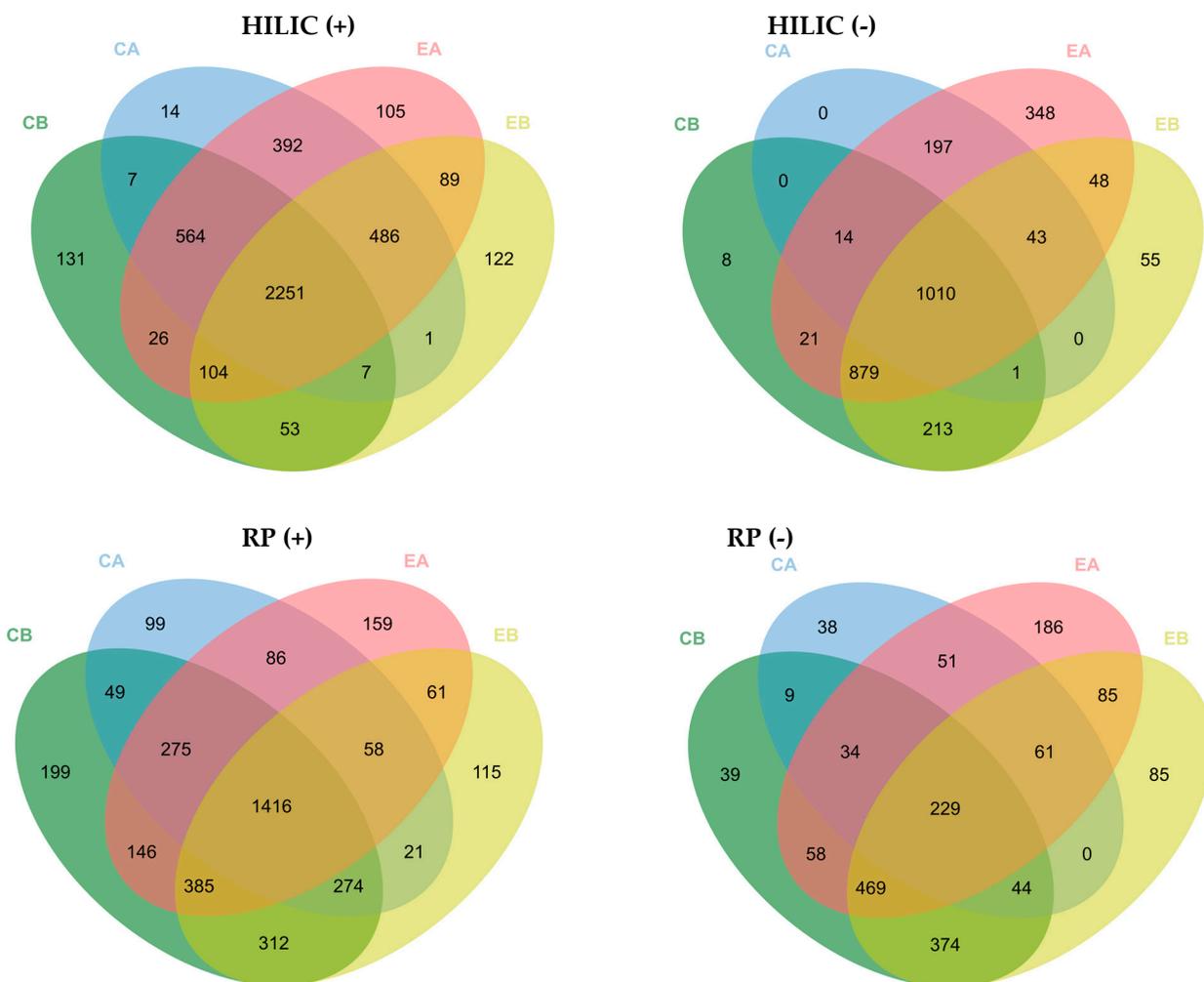
**Figure S2:** Box plots of raw  $\text{log}_2$  abundance values for HILIC (-) dataset which had a systematic shift in quantitative values (a) due to experimental variation (i.e. instrumentation, ionization efficiency, extraction efficiency). Plot (b) shows how the normalization procedure removes this systematic error.



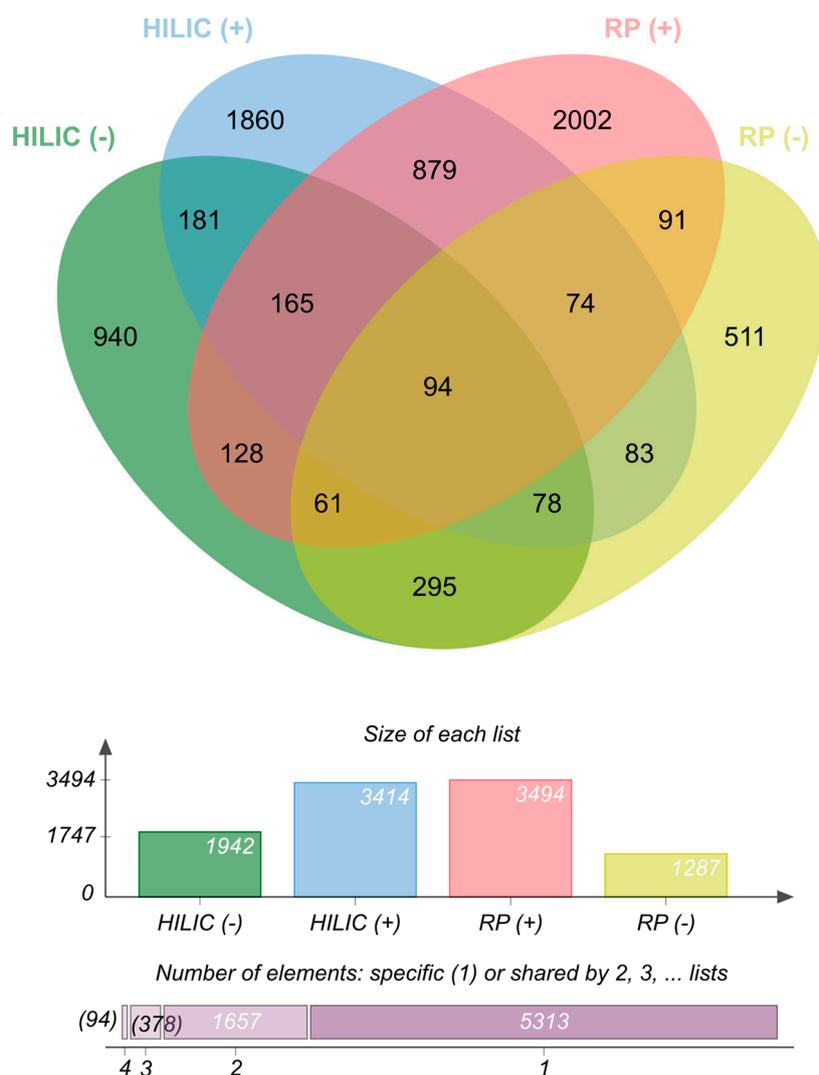
**Figure S3:** PCA plots of raw log<sub>2</sub> abundance values from the blank, controls, and samples before normalization, imputation, and filtering procedures.



**Figure S4:** (a) Histogram of the frequency of observations for each aligned peak (RT, MS<sup>1</sup>, and MS<sup>2</sup>) across the entire dataset (all 4 cores), including blanks and controls (55 total runs), before any data quality filtering steps and (b) a histogram of the HQFs across the 36 samples, after filtering out zeros, duplicates, and signals that were observed in the blanks or controls.



**Figure S5:** Venn diagrams showing the overlap of features between cores for each LC/MS condition. C = *Carex aquatilis*, E = *Eriophorum angustifolium*, A = Site A or low-centered polygon (LCP), and B = Site B or high-centered polygon (HCP).



**Figure S6:** Venn diagram (top) showing the overlap of HQFs between LC/MS conditions across all four cores obtained by calculating the neutral molecule from the MS<sup>1</sup> data, [M+H]<sup>+</sup> and [M-H]<sup>-</sup> ions (+/- 0.005 Da), excluding isomers and isobars. Bar graphs (bottom) show the total number of unique HQFs observed by each LC/MS condition and the number of LMW DOM features that were observed only once or multiple times across the four conditions.

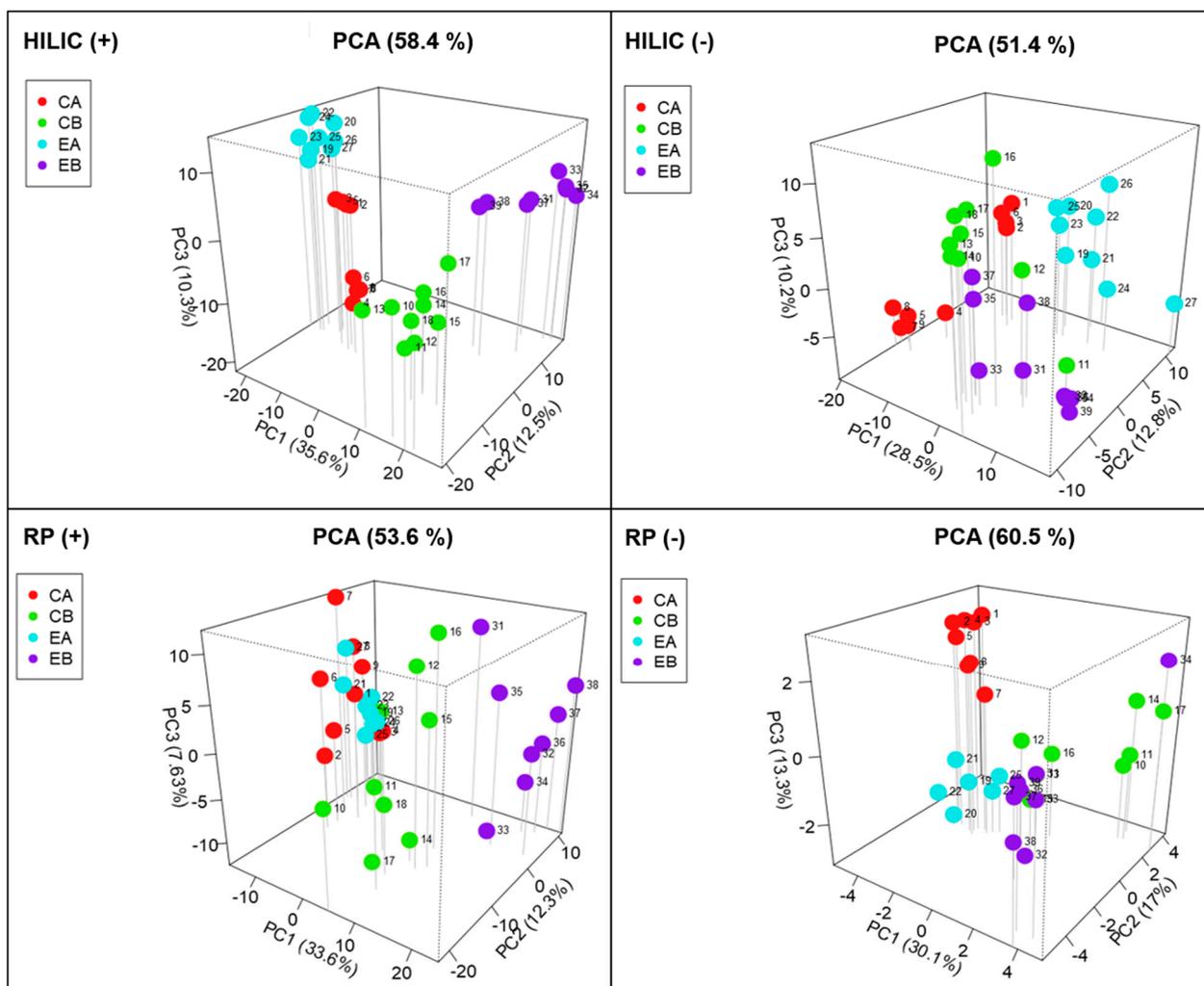


Figure S7: PCAs of HQFs by LC/MS condition.

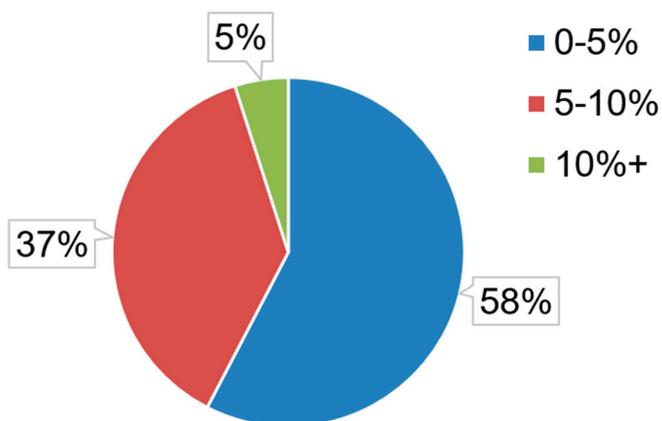
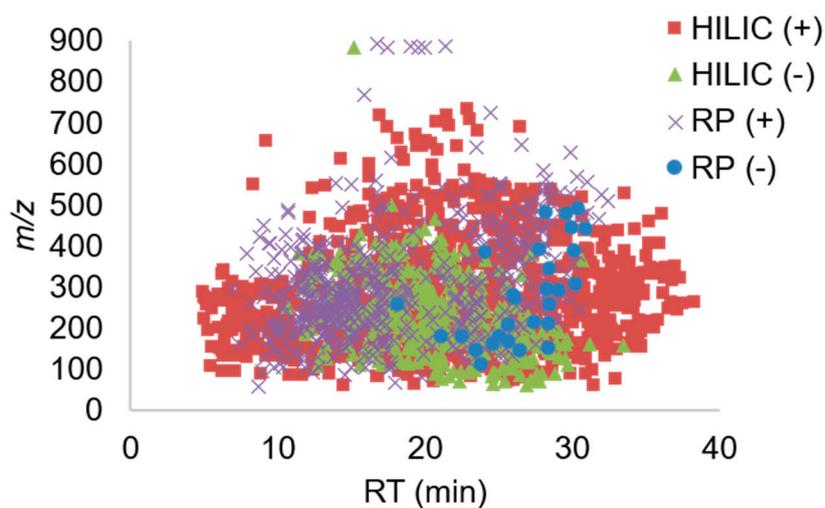
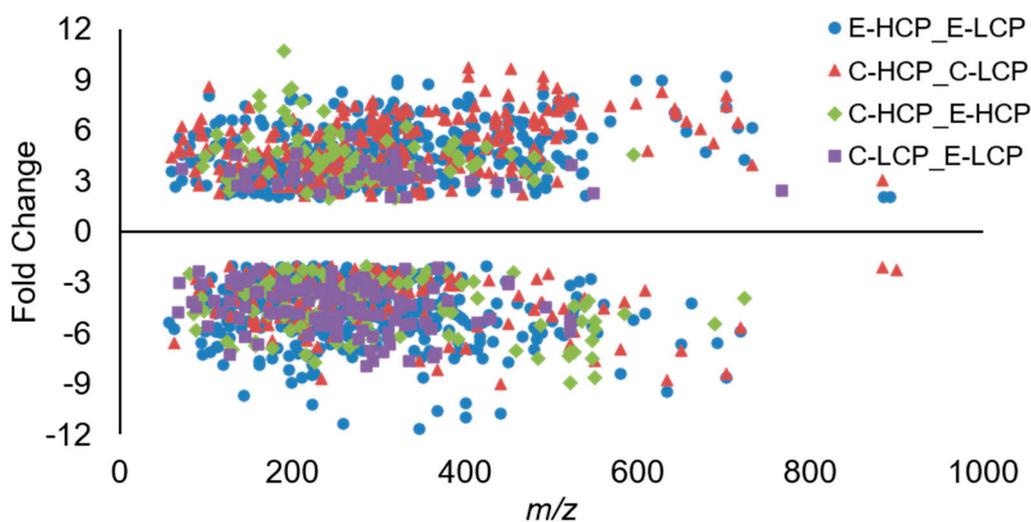


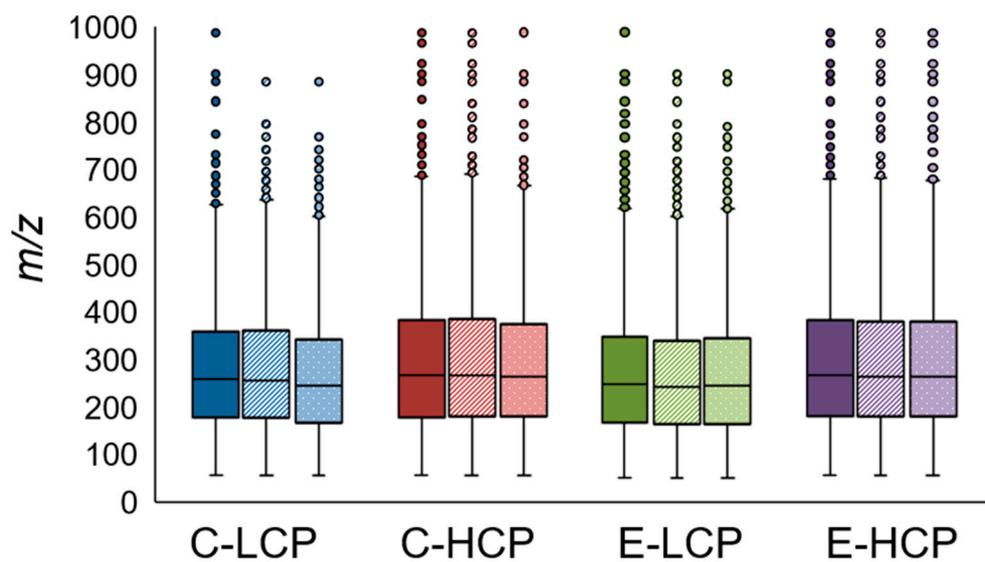
Figure S8: Pie chart with the results from a coefficient of variance analysis for peak areas of the differentially abundant LMW DOM features.



**Figure S9:** Distribution of molecular weight ( $m/z$ ) and retention time (RT) for differentially abundant features, detected across the 36 extracts, separated by LC/MS condition.



**Figure S10:** Distribution of molecular weight ( $m/z$ ) against fold change for LMW DOM features that were differentially-abundant due to polygon (E-HCP\_E-LCP and C-HCP\_C-LCP) or vegetation type (C-HCP\_E-HCP and C-LCP\_E-LCP).



**Figure S11:** Distribution of  $m/z$  by core and depth; solid color = top, stripes = middle, dots = bottom, C = *Carex aquatilis*, E = *Eriophorum angustifolium*

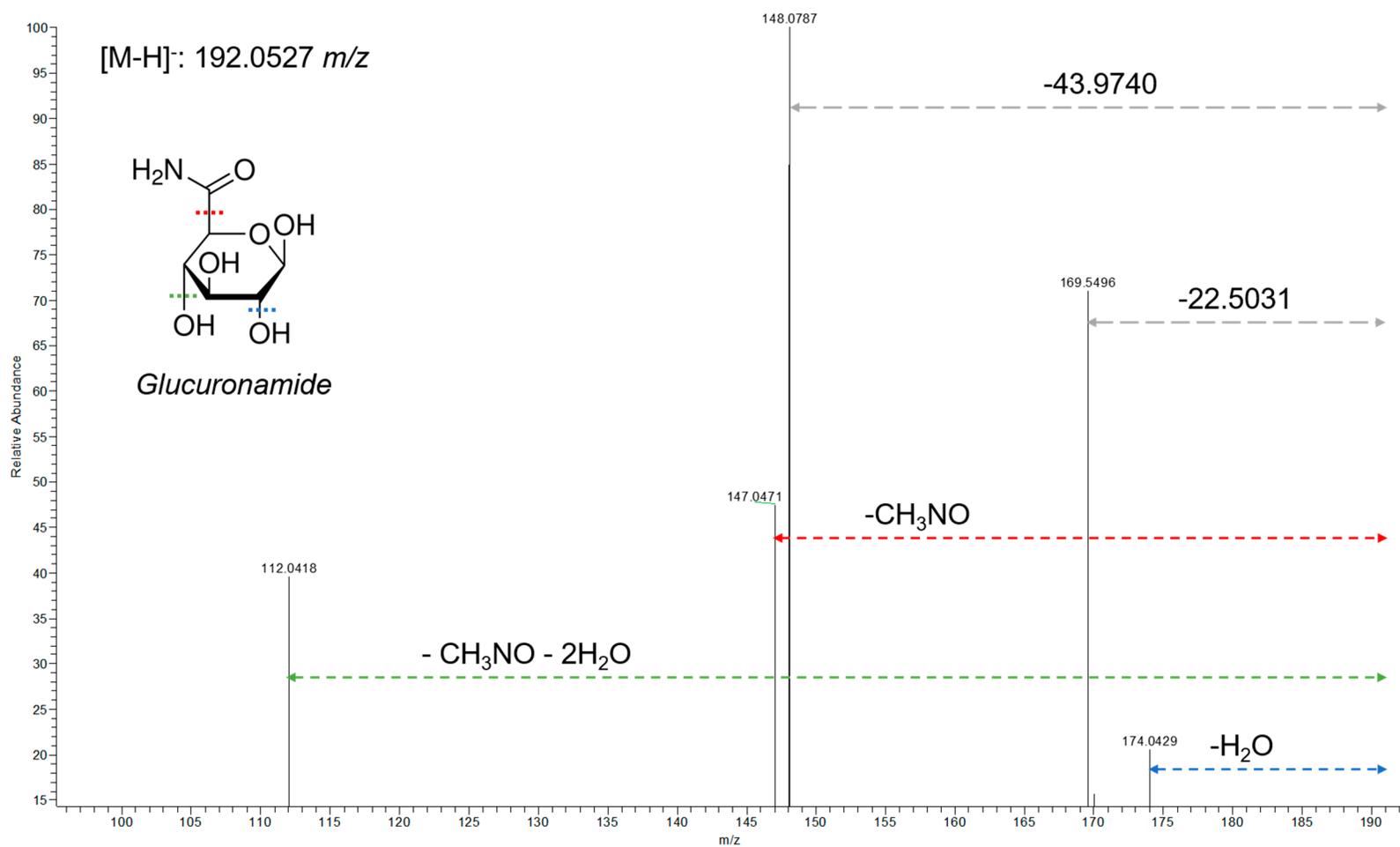


Figure S12: Fragmentation spectrum of  $[M-H]^-$  ion at  $m/z$  192.0527 showing characteristic neutral losses used for putative annotation.

## References

1. Sloan, V.L.; Liebig, J.; Hahn, M.; Curtis, B.; Brooks, J.; Rogers, A.; Iversen, C.; Norby, R. Soil temperature, soil moisture, and thaw depth, Barrow, Alaska Ver. 1. Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee, TN, USA: Next Generation Ecosystem Experiments Arctic Data Collection, 2014; <https://doi.org/10.5440/1121134>.
2. NOAA. National Overview. Available online: <http://www.ncdc.noaa.gov/sotc/national/2014/5> (accessed on August 12, 2014).
3. Chowdhury, T.R.; Herndon, E.M.; Phelps, T.J.; Elias, D.A.; Gu, B.H.; Liang, L.Y.; Wullschlegel, S.D.; Graham, D.E. Stoichiometry and temperature sensitivity of methanogenesis and CO<sub>2</sub> production from saturated polygonal tundra in Barrow, Alaska. *Glob. Chang. Biol.* **2015**, *21*, 722-737, doi:10.1111/gcb.12762.
4. Olivon, F.; Grelier, G.; Roussi, F.; Litaudon, M.; Touboul, D. MZmine 2 Data-Preprocessing To Enhance Molecular Networking Reliability. *Anal. Chem.* **2017**, *89*, 7836-7840, doi:10.1021/acs.analchem.7b01563.
5. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **2010**, *11*, doi:10.1186/1471-2105-11-395.
6. Nordstrom, A.; O'Maille, G.; Qin, C.; Siuzdak, G. Nonlinear data alignment for UPLC-MS and HPLC-MS based metabolomics: Quantitative analysis of endogenous and exogenous metabolites in human serum. *Anal. Chem.* **2006**, *78*, 3289-3295, doi:10.1021/ac060245f.
7. De Livera, A.M.; Sysi-Aho, M.; Jacob, L.; Gagnon-Bartsch, J.A.; Castillo, S.; Simpson, J.A.; Speed, T.P. Statistical Methods for Handling Unwanted Variation in Metabolomics Data. *Anal. Chem.* **2015**, *87*, 3606-3615, doi:10.1021/ac502439y.
8. Polpitiya, A.D.Q.; W.J.; Jaitly, N.; Petyuk, V.A.; Adkins, J.N.; Camp, D.G. 2nd; Anderson, G.A.S., R.D. DanTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics* **2008**, *24*, 1556-1558.
9. Evans, A.M.; DeHaven, C.D.; Barrett, T.; Mitchell, M.; Milgram, E. Integrated, Nontargeted Ultrahigh Performance Liquid Chromatography/Electrospray Ionization Tandem Mass Spectrometry Platform for the Identification and Relative Quantification of the Small-Molecule Complement of Biological Systems. *Anal. Chem.* **2009**, *81*, 6656-6667, doi:10.1021/ac901536h.
10. Zhou, B.; Xiao, J.F.; Tuli, L.; Ransom, H.W. LC-MS-based metabolomics. *Mol. Biosyst.* **2012**, *8*, 470-481, doi:10.1039/c1mb05350g.
11. De Livera, A.M.; Dias, D.A.; De Souza, D.; Rupasinghe, T.; Pyke, J.; Tull, D.; Roessner, U.; McConville, M.; Speed, T.P. Normalizing and Integrating Metabolomics Data. *Anal. Chem.* **2012**, *84*, 10768-10776, doi:10.1021/ac302748b.
12. Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M.Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods* **2016**, *13*, 731-740, doi:10.1038/nmeth.3901.
13. Barnes, S.; Benton, H.P.; Casazza, K.; Cooper, S.J.; Cui, X.; Du, X.; Engler, J.; Kabarowski, J.H.; Li, S.; Pathmasiri, W., et al. Training in metabolomics research. II. Processing and statistical analysis of metabolomics data, metabolite identification, pathway analysis, applications of metabolomics and its future. *J. Mass Spectrom.* **2016**, *51*, 535-548, doi:10.1002/jms.3780.
14. Kalivodová, A.; Hron, K.; Filzmoser, P.; Najdekr, L.; Janečková, H.; Adam, T. PLS-DA for compositional data with application to metabolomics. *J. Chemom.* **2015**, *29*, 21-28, doi:10.1002/cem.2657.
15. Jones, E., Oliphant, T., Peterson, P. . {SciPy}: Open source scientific tools for {Python}. **2014**.
16. Zhou B, W.J., Ransom HW. MetaboSearch: Tool for mass-based metabolite identification using multiple databases. *PLoS One* **2012**, *7*, doi:10.1371/journal.pone.0040096.PMID:22768229.
17. Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **2017**, *45*, D353-d361, doi:10.1093/nar/gkw1092.
18. Smith, C.A.; O'Maille, G.; Want, E.J.; Qin, C.; Trauger, S.A.; Brandon, T.R.; Custodio, D.E.; Abagyan, R.; Siuzdak, G. METLIN - A metabolite mass spectral database. *Ther. Drug Monit.* **2005**, *27*, 747-751, doi:10.1097/01.ftd.0000179845.53213.39.
19. Cui, Q.; Lewis, I.A.; Hegeman, A.D.; Anderson, M.E.; Li, J.; Schulte, C.F.; Westler, W.M.; Eghbalnia, H.R.; Sussman, M.R.; Markley, J.L. Metabolite identification via the Madison Metabolomics Consortium Database. *Nat. Biotechnol.* **2008**, *26*, 162-164, doi:10.1038/nbt0208-162.

20. Kim, S.; Thiessen, P.A.; Bolton, E.E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.Y.; He, J.E.; He, S.Q.; Shoemaker, B.A., et al. PubChem Substance and Compound databases. *Nucleic Acids Res.* **2016**, *44*, D1202-D1213, doi:10.1093/nar/gkv951.
21. Wishart, D.S.; Jewison, T.; Guo, A.C.; Wilson, M.; Knox, C.; Liu, Y.F.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E., et al. HMDB 3.0-The Human Metabolome Database in 2013. *Nucleic Acids Res.* **2013**, *41*, D801-D807, doi:10.1093/nar/gks1065.
22. Fahy, E.; Sud, M.; Cotter, D.; Subramaniam, S. LIPID MAPS online tools for lipid research. *Nucleic Acids Res.* **2007**, *35*, W606-W612, doi:10.1093/nar/gkm324.
23. Schlapfer, P.; Zhang, P.; Wang, C.; Kim, T.; Banf, M.; Chae, L.; Dreher, K.; Chavali, A.K.; Nilo-Pyanco, R.; Bernard, T.; Kahn, D.; Rhee, S.Y. Genome-Wide Prediction of Metabolic Enzymes, Pathways, and Gene Clusters in Plants. *Plant Physiol.* **2017**, *173*, 2041-2059.
24. Kind, T.; Tsugawa, H.; Cajka, T.; Ma, Y.; Lai, Z.J.; Mehta, S.S.; Wohlgemuth, G.; Barupal, D.K.; Showalter, M.R.; Arita, M., et al. Identification of small molecules using accurate mass MS/MS search. *Mass Spectrom. Rev.* **2018**, *37*, 513-532, doi:10.1002/mas.21535.
25. Koch, B.P.; Dittmar, T.; Witt, M.; Kattner, G. Fundamentals of molecular formula assignment to ultrahigh resolution mass data of natural organic matter. *Anal. Chem.* **2007**, *79*, 1758-1763, doi:10.1021/ac061949s.
26. Kind, T.; Fiehn, O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics* **2007**, *8*, doi:10.1186/1471-2105-8-105.
27. Kujawinski, E.B.; Behn, M.D. Automated analysis of electrospray ionization Fourier transform ion cyclotron resonance mass spectra of natural organic matter. *Anal. Chem.* **2006**, *78*, 4363-4373, doi:10.1021/ac0600306.
28. Ohno, T.; Parr, T.B.; Gruselle, M.C.I.; Fernandez, I.J.; Sleighter, R.L.; Hatcher, P.G. Molecular Composition and Biodegradability of Soil Organic Matter: A Case Study Comparing Two New England Forest Types. *Environ. Sci. Technol.* **2014**, *48*, 7229-7236, doi:10.1021/es405570c.
29. Antony, R.; Grannas, A.; Willoughby, A.; Sleighter, R.; Thamban, M.; Hatcher, P. Origin and sources of dissolved organic matter in snow on the East Antarctic ice sheet. *Environ. Sci. Technol.* **2014**, *48*, 6151-6159.
30. Tfaily, M.M.; Wilson, R.M.; Cooper, W.T.; Kostka, J.E.; Hanson, P.; Chanton, J.P. Vertical Stratification of Peat Pore Water Dissolved Organic Matter Composition in a Peat Bog in Northern Minnesota. *Journal of Geophysical Research-Biogeosciences* **2018**, *123*, 479-494, doi:10.1002/2017jg004007.