Characterization of dissolved organic matter in Lake Superior and its watershed using ultrahigh resolution mass spectrometry

Elizabeth C. Minor [a*1], Carla J. Steinbring [a1], Krista Longnecker [b,] Elizabeth B. Kujawinski[b]

[a] *Large Lakes Observatory and Dept of Chemistry and Biochemistry, University of Minnesota Duluth, Duluth MN 55811, USA*
[b] *Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole MA 02543, USA*

*Corresponding author.
E mail address: eminor@d.umn.edu (Elizabeth Minor).

[1] These authors contributed equally to the manuscript and should be considered co-first authors.

ABSTRACT

With the advent of ultrahigh resolution mass spectrometry, recent studies have begun to resolve molecular-level relationships between terrestrial and aquatic dissolved organic matter (DOM) in rivers, estuaries, mangrove swamps and their receiving oceans and lakes. Here, we extend ultrahigh resolution mass spectrometry techniques to Lake Superior, the largest freshwater lake in the world by area. Solid-phase extracted samples from the western arm of the lake and its watershed, including swamp, creek, river, lake-river confluence and offshore lake sites were compared using electrospray ionization (ESI) Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS). Results were analyzed using cluster analysis and van Krevelen diagrams. Chemical similarity appears related to hydrological proximity, terrestrial impact and flow conditions. For example, higher and lower flow samples from the same stream differ from one another. Toivola Swamp, Lake Superior, and the south shore river have diverse arrays of unique molecular formulae.relative to the north shore river and stream sampled in this data set. Lake Superior's unique elemental formulae, relative to its watershed samples, are primarily in the lignin-like and reduced hydrocarbon regions of van Krevelen diagrams. ESI-amenable Lake Superior DOM also has a higher proportion of formulae containing nitrogen or sulfur relative to the other samples. The degree of overlap among formulae within our data set is consistent with previous ESI FT-ICR-MS characterization of

terrestrial, estuarine and marine OM. There appears to be a conserved portion of formulae across natural OM samples, perhaps because these compounds are intrinsically refractory or because they are commonly generated as products of natural reworking processes.

Keywords
Dissolved organic matter
Natural organic matter
Electrospray ionization
Fourier-transform ion cyclotron resonance mass spectrometry
Ultrahigh resolution mass spectrometry
Lake Superior
Van Krevelen diagram
Cluster analysis
Lakes

1. Introduction

Dissolved organic matter (DOM) plays many important biogeochemical/ecological roles in marine and lacustrine water columns. It interacts with trace metals and anthropogenic organic molecules, affecting their solubility and bioavailability (e.g. Shiller et al., 2006). It is a photochemical reactant and a major food source for aquatic organisms (e.g. Pomeroy et al., 1979; Mopper et al., 1991). In many brown water systems, as well as in clear oligotrophic lakes such as Lake Superior, DOM absorbs most of the sunlight (e.g. Scully and Lean, 1994). Its chemical composition (along with its concentration in a system) determines its effectiveness in these biogeochemical roles. As a result, its characterization is an active area of research.

Until recently, molecular-level characterization of DOM relied on extensive wet chemical processing and traditional geochemical analyses. These approaches are applicable to a frustratingly small proportion of the dissolved organic carbon (DOC) in the water column (1% to 11% of marine DOC, as reviewed by Benner et al., 2002). Recent analytical chemistry advances, especially in the area of ultrahigh resolution mass spectrometry, have improved this situation. DOM studies employing ultrahigh resolution Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS), usually coupled with electrospray ionization (ESI), are still limited to

78 samples or DOM isolates with a low salt/OM ratio and to readily ionized
79 fractions of the extremely heterogeneous DOM pool. Nonetheless, they have
80 extended molecular-level characterization much further than any previous
81 technique as shown in multiple studies: river-to-ocean transects of water
82 column samples (Sleighter et al., 2008; Kujawinski et al., 2009), comparisons
83 of primarily terrigenous DOM with primarily marine water-column DOM
84 (Koch et al., 2005; Koprivnjak et al., 2009), comparisons of riverine, marine
85 and continental shelf porewater DOM (Schmidt et al., 2009) and studies of
86 bio- and photo- degradation of natural OM samples (e.g. Kim et al., 2006;
87 Kujawinski et al., 2004).

88 For most natural water studies, a sample concentration and clean up
89 approach is applied prior to FT-ICR-MS; the approach is usually either solid
90 phase extraction (C_{18} or PPL resins; e.g. Sleighter and Hatcher, 2008; Koch et
91 al., 2005; Schmidt et al., 2009) or reverse osmosis/electrodialysis (Koprivnjak
92 et al., 2009). Such sample pre-treatments collect a far greater proportion of
93 DOC than the more traditional geochemical methods of solvent extraction
94 and acid hydrolysis, with 25 to 62% of the DOC obtained by solid phase
95 extraction (Koch et al., 2005; Dittmar et al., 2008) and an average of 75% by
96 reverse osmosis/electrodialysis, Koprivnjak et al., 2009).

97 This study uses ESI FT-ICR-MS to characterize solid-phase extracted
98 DOM from Lake Superior, the largest freshwater lake by area in the world
99 (Herdendorf, 1982). It has a surface area of 8.2×10^{10} m^2 and exhibits a

100 biogeochemistry similar to oceans (Cotner et al., 2004). It is oligotrophic with
101 annual production of 65 g C m^{-2} yr^{-1}; Guildford and Hecky, 2000), an average
102 particulate OC concentration of ca. 0.08 mg C l^{-1} and DOC concentrations
103 ranging from 0.8 to 3.2 mg C l^{-1} (Urban et al., 2005). In situ primary
104 production is the main OC input into the system (Ostrom et al., 1998; Cotner
105 et al., 2004); OC from tributary streams represents ca. 10% of the total
106 annual photoautotrophic production in the lake (Cotner et al., 2004).
107 However, based upon C/N ratios and seasonal variations in DOC
108 concentration, terrigenous organic matter constitutes the bulk (80-95%) of
109 the DOC present in the Lake Superior water column, due most likely to
110 variations in residence time among the DOC pools (Urban et al., 2005). The
111 major sink for OC is through respiration of DOC by heterotrophic bacteria
112 (Cotner et al., 2004), although photodegradation may be another process that
113 alters DOM characteristics (Minor and Stephens, 2008; Biddanda and
114 Cotner, 2003). Less than 5% of the OC that sinks from the surface waters
115 accumulates in Lake Superior sediments (Baker et al., 1991). The carbon
116 cycle for the lake is not well understood, as estimates indicate that
117 approximately twice as much OC is lost per year, mainly through
118 remineralization but also through burial and outflow, than is added annually
119 by in situ primary production and input from rivers and precipitation (Cotner
120 et al., 2004; Urban et al., 2005).

121 Lake Superior, in contrast to the other Laurentian Great Lakes, has a
122 relatively long water residence time of 178 years (Quinn, 1992). Due to its
123 fresh water and its temperate location, the lake is dimictic, exhibiting
124 complete water column mixing twice a year. Estimates of DOC residence
125 time in the lake (approximately 8 years, Urban et al., 2005; 26 to 36 years,
126 Cotner et al., 2004) are considerably longer than water column vertical and
127 horizontal mixing rates (J. Austin, personal communication). The
128 combination of these features leads to an integration of the seasonally- and
129 geographically-varying delivery of organic matter from the atmosphere,
130 rivers, and streams.
131 The FT-ICR-MS study described here was designed as an investigation
132 of the variability in organic matter composition encompassed in the Lake
133 Superior watershed. It thus provides unique information useful to ecosystem
134 studies of Lake Superior itself and for comparison with similar
135 characterization of DOM from marine, estuarine and freshwater lotic systems
136 (e.g. Koch et al., 2005; Sleighter and Hatcher, 2008; Kim et al., 2006). Solid
137 phase (C_{18}) extraction was used to isolate the DOM from several sites within
138 the lake watershed, including a swamp, a creek (at higher and lower flow),
139 two rivers, a lake-river confluence site and an offshore lake site. The isolated
140 natural OM and its variation from sample to sample were then assessed.
141
142

2. Materials and methods

2.1. Study sites and sampling

Water samples were collected from the western arm of Lake Superior, which has been shown to have higher chlorophyll concentrations and greater productivity than the central and eastern regions of the lake (IJC Report, 1977). Seven samples were taken from six sites (Fig. 1, Table 1) throughout the western basin watershed. Toivola Swamp is a large blackwater swamp dominated by black spruce and ericaceous shrubs (Bridgham et al., 1998). St. Louis River is a fourth-order brown river containing high amounts of chromophoric material derived from the surrounding wetlands (including Toivola Swamp). Canal Park is within the confluence of the St. Louis River and Lake Superior, and, like an estuary in salt water systems, is subject to time-varying mixing of river and receiving-basin waters, though primarily affected by seiches rather than by tides. The Canal Park sampling site is located in the Duluth Harbor Basin and includes anthropogenic influences from shipping, industrial and urban activities. The mid-water column sample was taken from the lake at 50m depth at an offshore site (47º9'27"N, 91º17'19"W) with a water depth >300m. Amity Creek, which runs through rural/suburban Duluth and into the lake, is subject to flash flooding following storm events and was sampled during intermediate ("Amity Creek Higher Flow") and low flow ("Amity Creek Low Flow") periods to evaluate possible changes in OM delivered to Lake Superior under different flow conditions.

165 Brule River is a south-shore stream that drains different geological
166 formations and land use regimes than north-shore streams such as Amity
167 Creek (Detenbeck et al. 2004); the Brule River was sampled near its entrance
168 to the lake.

169 With the exception of the lake site, all water samples were collected
170 from the surface. Samples accessible by land were collected with acid- and
171 deionized water-rinsed steel or polypropylene buckets. The offshore lake
172 sample was collected aboard the R/V *Blue Heron* using Niskin bottles on a
173 CTD rosette. Samples were transported to the lab after collection, filtered
174 through 0.1-µm or 0.2-µm Whatman Polycap cartridge filters (Whatman Inc.;
175 Florham Park, New Jersey, USA), and stored at 4 °C until further processing.
176 Prior to use, each cartridge filter was flushed with deionized water for >30
177 min and rinsed briefly with a small volume of sample water, which was then
178 discarded. Previous testing has found this treatment to yield low-carbon
179 blanks (Minor and Stephens, 2008; Kruger et al., 2011).

180 *2.2. Sample extraction*

181 Aliquots (1 l) of the filtered water samples and a MilliQ water blank
182 were acidified to pH 2 using 6M HCl, extracted with solid-phase C_{18} resin
183 (3M Empore disks) and eluted with methanol (MeOH) and MilliQ water
184 (90:10 v/v) following the procedures of Minor and Stephens (2008). The
185 resulting eluent was dried in a vacuum oven at ca. 40 °C and stored in the
186 dark until analysis with ESI FT-ICR-MS.

187 Recovery (extract content/initial sample content) during extraction of
188 samples was determined by analyzing aliquots of extract and initial sample
189 using UV-visible spectrophotometry to quantify colored DOM and DOC
190 analysis to measure carbon concentration. Recovery calculations were
191 performed as in Simjouw et al. (2005). For DOC analysis, selected extracts
192 were dried in a vacuum oven (to remove MeOH from the eluent) and
193 resuspended in a known volume of MilliQ water, from which an aliquot was
194 taken for analysis. Initial samples and extract aliquots were acidified to pH
195 2 (using 6M HCl) prior to measurement as non-purgeable OM on a Shimadzu
196 high temperature combustion TOC analyzer (TOC_{VSH}). The recovery data are
197 shown in Table 1.

198 *2.3 Ultrahigh resolution mass spectrometry*

199 Before FT-ICR-MS analysis, the dry samples were dissolved in
200 ultrapure water at a concentration of 1 mg/ml. A 1 to 4 dilution of each
201 sample was prepared in 50:50 MeOH:water for analysis in the negative ion
202 mode, and in 50:50 MeOH:water plus 0.1% HCOOH for analysis in positive
203 ion mode. These dilutions were prepared immediately before analysis to limit
204 ester formation (McIntyre and McRae, 2005). A Suwannee River Fulvic Acid
205 (SRFA) sample (International Humic Substances Society) was prepared in
206 the same fashion for analysis at a range of dilutions to evaluate the effect of
207 concentration upon peak identification and intensity and to determine the
208 appropriate dilution range for our samples. This SRFA sample, a

purchasable reference material isolated from the Suwannee River in Georgia, was also used as an external standard to ensure similar instrument performance across analysis periods.

Ultrahigh resolution mass spectrometry was performed with a Thermo Scientific 7 Tesla electrospray ionization Fourier transform ion cyclotron resonance mass spectrometer located at the Woods Hole Oceanographic Institution. All samples were run with instrument settings optimized by tuning on the SRFA standard. The instrument was externally calibrated weekly with a standard solution from Thermo Fisher Scientific, which calibrates to a mass accuracy of <2 ppm. A scan range of m/z 150-1000 was used and 200 scans were collected in each run. The average resolving power was 400,000 at m/z 400, where resolving power is calculated as the m/z value divided by the peak width at 50% peak height ($M/\Delta M_{50\%}$; Marshall et al., 1998). Scans were co-added using MIDAS (Modular ICR Data Acquisition and Analysis, version 3.2, from the National High Magnetic Field Laboratory, Tallahassee FL), zero-filled once, Hanning apodized and fast Fourier transformed. The reproducibility of DOM samples using ESI FT-ICR-MS has only recently been addressed (Kido Soule et al., 2010). Therefore, to confirm reproducibility, data acquisition was repeated in triplicate for the Canal Park and Amity Creek Low Flow samples. In addition, the SRFA standard was used to compare samples analyzed in March 2009 and March 2010.

231 In addition to external calibration, an internal re-calibration was
232 applied to the aligned peak list (using MIDAS) prior to final peak
233 assignment. Three sets of compounds differing only by CH_2 groups and one
234 additional compound differing in degree of unsaturation from the others were
235 chosen as internal calibrants based on their presence in all samples, low error
236 and high average peak intensity (see Supplementary Table 1).

237 *2.4. Formula assignment*

238 For each spectrum, automated analysis was used to assign chemical
239 formulae to peaks with a signal to noise ratio (S/N) >3. To see how much this
240 choice of S/N affected our results, we also performed the same processing
241 with S/N >5 (see Fig. 2 and Table 2). As relative trends among samples
242 remained similar at both S/N levels (Fig. 2) and a considerable amount of
243 differentiating sample information may be lost when S/N>5 (Fig. 2 and Table
244 2), the data used for the remainder of the study are those processed at S/N>3.
245 Formula assignment was done using the Compound Identification Algorithm
246 program developed by Kujawinski and Behn (2006). Allowed elements were
247 ^{12}C, ^{1}H, ^{16}O, ^{14}N, ^{32}S, ^{31}P and ^{13}C. The maximum allowed formula error was 1
248 ppm, the relation error 20 ppm, and the mass limit for empirically assigning
249 elemental formulae was 500 Da. Formulae above 500 Da were assigned
250 through the detection of homologous series. Mass values with more than one
251 possible elemental formula were assigned according to selection rules that
252 had been optimized for terrestrial DOM (Kujawinski et al., 2009). If no

chemical formula matched an m/z value within the allowed error, the peak was not included in the list of elemental formulae.

Number-averaged elemental ratios were calculated by dividing the sum of the elemental ratios by the total number of assigned formulae for a given sample:.

$$\frac{H}{C}_n = \frac{\sum \frac{H}{C}_i}{\sum Peaks} \qquad (1)$$

The magnitude-averaged elemental ratios H/C_w, O/C_w, N/C_w, P/C_w, and S/C_w were also calculated (as in Sleighter and Hatcher, 2008) from the formula assignments within each sample, in this case scaled to peak response for each formula. For example, H/C_w was calculated using the following equation:

$$\left(\frac{H}{C}\right)_w = \frac{\Sigma\left(\frac{H}{C} * Magnitude\right)}{\Sigma Magnitude} \qquad (2)$$

where H and C represent the numbers of H and C atoms in each elemental formula and magnitudes are determined from peak height for each assigned formula's m/z value . Elemental percentages $\%F_S$, $\%F_N$, $\%F_P$, $\%F_{C13}$ were also determined for each sample as the percentage of formulae within a sample containing at least one atom of the specific element or isotope. This allowed insight into the distribution of molecules between sample sites, though it is

not directly analogous to elemental ratios performed upon bulk OM fractions. For additional information about molecular structure, number-averaged double bond equivalents (DBEs; rings plus double bonds) were calculated using the following formula:

$$DBE = \left(C + {}^{13}C\right) - \frac{H}{2} + \frac{N}{2} + 1 \quad (2)$$

The elemental formula percentages %CHO, %CHON, %CHOS, %CHONS, %CHOP, %CHONP, %CHONSP represent possible combinations of elements included in the compound identification algorithm. For each sample, the percent of assigned formulae containing only the given elements was calculated.

2.5. Multivariate analysis

Combinatorial cluster analysis with the group average method was performed using MATLAB and the Fathom toolbox from D. Jones at the University of Miami. The Bray-Curtis dissimilarity measure was used to calculate the distance matrix, and the presence or absence of peaks in each sample was used as the data input. The *x*-axis in the cluster diagram represents the proportion of initial information content remaining after each step.

2.6. Molecular formula graphical analysis

Van Krevelen diagrams (as in Kim et al., 2003) were constructed from elemental ratios to further investigate molecular-level variation among the

293 samples. While each major biogeochemical class of compounds yields a range
294 of characteristic H/C and O/C ratios that can be correlated with the diagram
295 (Kim et al., 2003), it should be emphasized that peaks falling within this
296 range are not definitively identified as such compounds. The van Krevelen
297 diagrams are based upon the elemental ratio in each identified formula and
298 (in contrast to Table 3) do not include relative peak height information.
299
300 **3. Results and discussion**
301 *3.1. Mass spectra*
302 In all positive ion mode ESI samples, a contamination signal appeared
303 with peaks spaced approximately 44 *m/z* values apart. This contamination
304 suppressed the DOM signal. Samples in negative ion mode yielded spectra
305 with a high relative abundance for DOM peaks; blank spectra were generally
306 low in peak abundance and amplitude, though there was a significant peak at
307 *m/z* 412.96643 in the process blank that also appeared in the watershed
308 samples. This peak has been identified as an artifact from extraction with
309 C_{18} disks [Sleighter and Hatcher (2008) report its *m/z* value as 412.96638].
310 Because of better sample response and low process-blank response, all data
311 presented here are based upon negative ion mode mass spectra.
312 The majority of the peaks were in the range *m/z* 250-600 (see Fig. 3).
313 Our results are similar to those presented previously in showing multiple
314 ions within integer *m/z* values (e.g. Brown and Rice, 2000; Kujawinski et al.,

315 2002). DOM peaks were singly charged, as determined using the isotopic
316 distribution of carbon (as shown by Stenson et al., 2002). For each sample,
317 greater than 2,000 peaks exceeded an S/N threshold of 3. Peaks in the blank
318 contributed between 21 and 87 peaks to this number, depending upon the
319 sample. As these blank peaks constituted <5% of each sample's peaks, and
320 with the exception of m/z 412.96643, had low peak amplitude, no further
321 action was taken concerning blank correction of the samples.
322 Negative ion mode ESI (vs. positive ion mode ESI) FT-ICR-MS has
323 been shown to generate more m/z peaks from natural OM samples (e.g.
324 Hertkorn et al., 2008). The better response of natural DOM to negative
325 ionization conditions may explain why we had good sample response and
326 little blank contribution in negative ion mode as compared to significant
327 blank/contamination contributions to our analyses in positive ion mode. It is
328 also likely that the contamination compounds preferentially ionize in positive
329 ion mode.
330 *3.2. Instrument reproducibility*
331 The data from the SRFA analyses in March 2009 and March 2010, as
332 well as replicate analyses of Canal Park and Amity Creek Low Flow samples,
333 were used to assess instrument reproducibility and repeatability. From the
334 aligned peak list, the proportion of peaks shared between replicates was used
335 as a measure of the degree of reproducibility of ESI FT-ICR-MS analysis of
336 these samples (Table 2). The mass spectra from all the sample sites in the

337 lake watershed share 22% of peaks at S/N>3 and 60% of peaks at S/N>5. The
338 Canal Park and Amity Creek Low Flow replicate spectra share 65% and 64%
339 of peaks, respectively, at S/N>3 (87 and 92%, respectively, at S/N>5). The
340 two SRFA runs had peak overlap of 62% at S/N>3 and 91% at S/N>5. This
341 reproducibility is similar to that described previously for this instrument
342 (Kido Soule et al., 2010).

343 The reproducibility of our ESI FT-ICR-MS method was also shown by
344 comparing within-replicate to between-sample elemental formula
345 assignments (Table 3). The average and sample standard deviation of
346 elemental ratio and percentage formula data in Table 3 were calculated for
347 triplicate measurements of Canal Park and Amity Creek Low Flow samples,
348 indicating the variability of the instrument and formula assignment. The
349 resulting sample standard deviations indicate that, for most elemental
350 formula comparisons, the differences between sites were greater than within
351 sample variability.

352 The percentage of peaks shared by the lake watershed samples (22% at
353 S/N>3; 60% at S/N>5) was similar to or higher than the percentage of peaks
354 shared by solid phase extraction (SPE) extracts from swamp, offshore coastal
355 water and mid-river water in a study of the lower Chesapeake Bay (29% with
356 negative ion mode at S/N=5; Sleighter and Hatcher, 2008). While shared
357 formula assignments do not necessarily mean common chemical structures,
358 they do indicate that a portion of the DOM may be conserved across aquatic

359 regimes, either due to inherent refractoriness or because it is generated
360 through reworking of OM from different sources (as hypothesized by
361 Reemtsma et al., 2008).

362 *3.3. Cluster analysis*

363 Classification via cluster analysis (using peak presence vs absence
364 data) was used to investigate chemical similarity among samples and
365 replicates (Fig. 2). Regardless of S/N criterion chosen, replicate analyses
366 generally group together. The St. Louis River sample does group with the
367 Canal Park replicates in Figure 2A, which is not too surprising as the St.
368 Louis River sample is upstream of, and directly hydrologically connected to
369 the Canal Park site. The editing of information content in the more stringent
370 S/N>5 case leads to a more pronounced split in the grouping for Canal Park
371 replicate analyses and the loss of clear differentiation between St. Louis
372 River and Canal Park, which are in hydrological proximity, and Amity Creek,
373 which is a north-shore stream approximately 10 km away. At both S/N
374 levels, Toivola Swamp and Lake Superior group separately from the tributary
375 samples, and the south-shore site, Brule River, is separated from the Amity
376 Creek, St. Louis River, and Canal Park sites, indicating a difference between
377 south-shore and north-shore DOM. With S/N>3 (Fig 2A), replicates group
378 more clearly and there is separation of SRFA, Lake Superior, Toivola Swamp,
379 south shore Brule River and Duluth area stream/rivers. We thus focus our
380 further results and discussion on the S/N>3 data.

381 The clustering of the Lake Superior tributary samples likely occurs
382 because the DOM has similar sources and processing in these ecosystems
383 with relatively similar water residence times, climate and land use patterns.
384 The stream/river sites all have relatively fresh OM that is primarily
385 terrestrially-derived, with shorter residence times than the 178-year
386 residence time of water in the lake (Quinn, 1992), where the OM is believed
387 to be primarily autochthonous (Cotner et al., 2004). The swamp sample is
388 most likely differentiated from the other Lake Superior watershed samples
389 by both differences in organic matter source (from the swamp's unique plant
390 material) and water flow patterns. While Suwannee River also represents a
391 terrestrially-derived source of OM, it appears to be considerably different
392 from the Lake Superior watershed samples, perhaps because its DOM had
393 been isolated using a different solid-phase extraction technique and also
394 because of the differences in primary vegetation in its watershed (in Georgia
395 and Florida) relative to northern Minnesota.
396 The Amity Creek samples were grouped on the basis of flow rate. The
397 high flow sample in Fig 2A is aligned more closely with the higher-order St
398 Louis River and Canal Park.
399 *3.4. Peak assignment*
400 The methods described in Section 2.4 were able to assign elemental
401 formulae to most of the peaks (>97%) with S/N>3 (Table 3).

Of all the lake watershed samples, Toivola Swamp has the highest number-averaged DBE and DBE normalized to the number of carbons (DBE/C), and lowest number-averaged and magnitude-averaged H/C ratios and intermediate number-averaged and magnitude-averaged O/C ratios; this combination indicates that it contains the most condensed OM and most likely includes the most significant contribution from aromatic compounds.

Lake Superior (mid-water column) has the next highest DBE and DBE/C, and appears to be the least oxidized (i.e. it has the lowest O/C_n and O/C_w ratios). It also has the highest number-averaged and magnitude-averaged N/C values in the data set, as would be expected if higher proportion of its organic matter were derived from phytoplankton and bacteria (Meyers 1994). Our O/C_w ratio for the lake (0.38) is very similar to values reported for whole water DOM from the Dismal Swamp in Virginia, USA (0.39; Sleighter and Hatcher, 2008), SPE-extracted DOM from a Chesapeake Bay sub-estuary (0.33-0.35; Sleighter and Hatcher 2008), and C_{18}-extracted DOM from Atlantic Ocean surface water (0.34-0.36; Kujawinski et al., 2009).

The greatest elemental ratio difference between Amity Creek Higher Flow and Low Flow samples is in H/C_n and H/C_w; the Higher Flow sample has lower number-averaged and magnitude-averaged H/C values and slightly higher DBE/C,perhaps indicating a shift toward somewhat more aromatic and terrestrially-derived material.

Across sampling sites, the P/C_w and S/C_w ratios are low and relatively invariant (ranging from 0.001 to 0.008 and 0.00 to 0.07, respectively) and number-averaged values are even lower. With the exception of Brule River, which has a higher ratio, and Toivola Swamp, which has a lower ratio, H/C_w values for our samples (Table 3) are also relatively invariant and similar to those reported by Sleighter and Hatcher (2008) for the Virginia swamp and upriver SPE-DOM samples (1.25 and 1.29) and somewhat lower than reported by Kujawinski et al. (2009) for Atlantic Ocean surface water (1.3-1.32).

Elemental formula proportions (%; Table 3) yield additional information about the compositional differences in DOM among sites. More than half of all the assigned compounds in each sample contain only C, H and O. The lake has the lowest proportion of exclusively CHO compounds (56%), in general lower than previously reported for surface ocean, deep ocean and river samples (72.2 to 93.7%; Kujawinski et al., 2009). However, river/creek samples with more terrestrial input have a higher proportion of CHO compounds (65 to 77%, Table 3) than our open lake site. The Amity Creek Low Flow sample contains a smaller proportion of CHO compounds (69%) than the Amity Creek Higher Flow sample (77%). Our swamp sample (64% CHO) fell between the lake and river/creek values.

The next largest abundance of compounds is exclusively CHON compounds (16-29%). The lake sample has the highest proportion (29%),

446 consistent with its highest proportion of formulae containing at least one N
447 atom (F_N 41%). For all sites, there is a higher abundance of compounds
448 containing one (non-oxygen) heteroatom than those with two or more
449 heteroatoms. The south-shore Brule River site has the highest proportion of
450 compounds containing CHONP (9.4%), closely followed by Lake Superior and
451 then Toivola Swamp. Lake Superior and Toivola Swamp have the highest
452 proportion of CHOS compounds (1.2% and 1.1%, respectively). The St. Louis
453 River and Canal Park sites share very similar proportions of compound types.
454 The heteroatom composition of extracted DOM can also be viewed in
455 terms of the proportion of identified formulae containing that heteroatom
456 (Table 3). For all the sampling sites the proportion of formulae containing N
457 (F_N) was greater than that for other non-oxygen heteroatoms (F_P, F_S) The
458 lake site had a higher proportion of both sulfur- and nitrogen-containing
459 formulae than the other sites (see Table 3, F_S and F_N values). The Brule
460 River, Lake Superior and Toivola Swamp had higher proportions of P-
461 containing formulae than the other sites (Table 3, F_P values).
462 The presence of sulfur compounds in the DOM samples (F_s ranges from
463 1.5% to 3.4%) is not surprising. S-containing organic compounds have been
464 identified using negative ionization ESI FT-ICR-MS of solid-phase extracted
465 DOM from the Weddell Sea (D'Andrilli et al., 2010) and the North Atlantic
466 Ocean (Kujawinski et al., 2009).

467 Lake Superior is oligotrophic and phosphorus-limited (Sterner et al.,
468 2004). It is surprising, therefore, in view of the extremely high nitrate:
469 phosphate ratios for Lake Superior water samples (Sterner et al., 2007), that
470 Lake Superior extracted DOM has a fairly high F_P of 11.4%. One might have
471 expected that under extremely P-depleted conditions, P-containing
472 compounds in the DOM would be very quickly taken up by biota and
473 reincorporated into the particulate pool.

474 *3.5. Molecular Formula graphical analysis*

475 Van Krevelen diagrams can help distinguish compound classes in
476 samples on the basis of the characteristic elemental ratios that each major
477 biogeochemical compound class generally exhibits (Sleighter and Hatcher,
478 2007); however, it should be emphasized that these elemental ranges are not
479 sufficient for providing unequivocal compound class identification
480 (Reemtsma, 2010). The van Krevelen diagrams for all of the Lake Superior
481 watershed samples were similar (Fig. 4), with a dominance of lignin-like
482 formulae. The dominance of lignin-like components (characterized by an O/C
483 molar ratio of ca. 0.25-0.5 and an H/C molar ratio of ca. 1-1.5) is similar to
484 that seen in swamp and offshore marine DOM (Sleighter and Hatcher, 2007).
485 This region potentially includes carboxyl rich alicyclic molecules (CRAM)
486 which would fall into the same van Krevelen space (Hertkorn et al., 2006).
487 Tannin-like formulae also appear in all samples, indicating an additional
488 terrigenous component to the DOM. There are also high numbers of

489 formulae corresponding to condensed hydrocarbons, which may indicate black
490 carbon from incomplete combustion of forest/grass or fossil fuels in the
491 watershed. Black carbon, while often thought of as particulate, has been
492 measured in the dissolved phase in other aquatic systems (Kim et al., 2004;
493 Dittmar, 2008). In contrast to the lignin-like and condensed hydrocarbon
494 formulae, there are considerably fewer lipid, carbohydrate, or protein
495 formulae present.

496 The Lake Superior and Brule River samples have a higher abundance
497 of lipid-, carbohydrate- and protein-like formulae than other sites (with the
498 exception of lipid-like formulae in Amity Creek Lower Flow). A previous
499 study using Fourier transform infrared spectroscopy (FTIR) has shown that
500 Lake Superior tributary sites contain higher molecular weight OM, rich in
501 lignin and protein compounds, while the open lake contains OM that appears
502 less enriched in protein and lignin moieties and more enriched in
503 carbohydrate and aliphatic material (Minor and Stephens, 2008). Our FT-
504 ICR-MS data (based upon peak presence-absence data) is consistent in terms
505 of carbohydrate and aliphatic trends but does not show the same protein
506 relationship.

507 Comparing the Amity Creek sites, the Low Flow sample has more
508 lipid-like, carbohydrate-like and protein-like formulae than the Higher Flow
509 sample, which is indicative of autochthonous (in-creek generated) OM.
510 During low flow, the Amity Creek site is likely to have a good mix of OM from

511 both in-creek and terrestrial origins. The creek's short water residence time
512 does not allow extensive reworking of terrestrially-derived DOM; however,
513 the lower flow does allow increased microbial activity and enhanced
514 autochthonous production of DOM.

515 While Figure 4 emphasizes the similarity of the compounds between
516 the sites, van Krevelen diagrams for the compounds unique to each sampling
517 site (Fig. 5) also allow us to see the differences in OM that are unique to each
518 site and contribute to its characteristic signature. Brule River, Lake
519 Superior, and Toivola Swamp have the highest proportions of unique peaks
520 (21%, 19%, and 14%, respectively, , S/N>3) in the dataset, supporting the
521 results from the cluster diagram (Fig. 2A). Of all the watershed samples, the
522 Brule River sample contains the most diverse and evenly-dispersed
523 arrangement of compound classes. The Lake Superior and Toiviola Swamp
524 sites have strong, but differing lignin-like and condensed hydrocarbon
525 contributions, as well as some unique carbohydrate/protein contributions.

526 These diagrams show that Lake Superior has an OM component that
527 is compositionally diverse and distinct from the organic matter found in its
528 tributaries. It is surprising that the unique component of its OM resembles
529 terrestrially-derived DOM (lignin-like and condensed hydrocarbon formulae),
530 but this may indicate reworking of terrestrially-derived material by microbial
531 and photochemical processes in this highly oligotrophic and clear water lake.
532 For example, condensed hydrocarbons make up a considerable proportion of

533 the unique formulae in Lake Superior. Kim et al. (2006) found that formulae
534 with low H/C values (black-carbon like) were enhanced upon biodegradation
535 of stream water. Previous FTIR analysis of Lake Superior DOM samples
536 showed that this oligotrophic lake contains a wide range of carbohydrate and
537 other allochthonous OM that appears modified from the terrestrial OM,
538 making it different on the molecular level (Minor and Stephens, 2008). The
539 high proportion of formulae unique to the lake site (Fig. 5) is consistent with
540 this, indicating that the OM in the lake is not simply the material delivered
541 by its tributaries but is modified within the lake itself.

542 The Amity Creek Low Flow sample exhibits an intermediate
543 proportion (8%) of unique compounds (including lignin and carbohydrate-like
544 compounds) while the Higher Flow sample has a lower proportion (2%).
545 Comparing just the higher and lower flow samples (at S/N>3), 63% of all the
546 peaks found in Amity Creek analyses are unique to the three replicates of the
547 Low Flow sample and do not appear in the Higher Flow sample; and 1% are
548 unique to the Higher Flow sample. These Amity Creek results, coupled with
549 the elemental trends given above, indicate that there is an increase in
550 autochthonous OM during low flow periods overlaying allochthonous OM
551 inputs occurring at both higher and lower flow periods.

552 The St. Louis River and Canal Park samples both contain very small
553 proportions of unique compounds. This is consistent with our cluster analysis
554 results (Fig. 2A), indicating that these sites are very similar, most likely

555 because of their hydrological proximity. Because of their similarity in DOM,
556 few peaks identified are solely unique to either of these sites.
557
558 **4. Conclusions**
559 Ultrahigh resolution mass spectrometry was used to characterize C_{18}-
560 extracted DOM within a large lake watershed. Molecular level
561 characterization in conjunction with cluster analysis was used to provide
562 insight into the similarities and differences among sites. Toivola Swamp,
563 Lake Superior, and the Brule River (the south shore river in our data set)
564 have diverse arrays of unique molecular formulae.relative to the north shore
565 river and stream. Mid-water column Lake Superior exhibits unique OM
566 components relative to river, creek and swamp sites in its watershed,
567 indicating that it contains a pool of OM that is different from the OM in its
568 tributaries. Lake Superior's unique formulae are primarily in the lignin-like
569 and condensed hydrocarbon regions of van Krevelen space, though its DOM
570 also has a higher proportion of heteroatom (N and S) containing formulae
571 than the other sites. Sites in hydrological proximity exhibit strikingly similar
572 molecular-level compositions relative to the rest of the data set. Comparison
573 of higher flow vs. low flow stream samples from Amity Creek shows that flow
574 influences the quality of the OM in the C_{18}-extractable DOM pool, with the
575 low flow sample containing more unique formulae than the higher flow
576 sample. While there are considerable differences in the DOM composition of

577 the various samples, 22% (at S/N>3) to 60% (at S/N>5) of the peaks were
578 shared by all the samples in the set. Sharing exact mass peaks does not
579 necessarily imply sharing the same compounds, as structural isomers by
580 definition exhibit the same elemental formulae. However, that 22 to 60% of
581 the peaks are shared may indicate that there is an intriguing component
582 conserved across highly variant biogeochemical regimes.
583 This is the first time ESI FT-ICR-MS has been applied to a large lake
584 as well as its selected tributaries to study variations in DOM composition.
585 Such an approach can provide insight into the spatial and temporal
586 variability in DOM composition in such watersheds and provide data for
587 comparison with other aquatic systems.
588
589 **Acknowledgements**
590 This study was supported in part by NSF grant OCE-0825600 (to E.C.M.).
591 C.S. was supported in part by a National Science Foundation Graduates in K-
592 12 Education grant to the University of Minnesota Duluth. Travel support
593 was provided to C.S. through the travel fund at the WHOI FT-MS facility
594 which was funded by the WHOI Director of Research and NSF grant OCE-
595 0751897. The authors thank M. Kido Soule for FT-ICR-MS assistance and the
596 funding sources of the WHOI FT-MS facility (NSF grant OCE-0619608 and
597 the Gordon and Betty Moore Foundation). Thanks also go to B. Stephens and

598 B. Kruger (Large Lakes Observatory) for sharing samples and to S. Stark
599 (UMD Geology) for assistance in generating Fig. 1.

600

601 **References**

602

603 Baker, J.E., Eisenreich, S.J., Eadie, B.J., 1991. Sediment trap fluxes and
604 benthic recycling of organic carbon, polycyclic aromatic hydrocarbons, and
605 polychlorobiphenyl congeners in Lake Superior. Environmental Science and
606 Technology 25, 500-109.

607

608 Benner, R. 2002. Chemical composition and reactivity. In: Hansell, D.A. and
609 Carlson, C.A. (Ed.), Biogeochemistry of Marine Dissolved Organic Matter.
610 Academic Press, San Diego, CA pp. 59–90.

611

612 Biddanda, B.A., Cotner, J.B., 2003. Enhancement of dissolved organic matter
613 bioavailability and its role in the carbon cycle of Lakes Superior and
614 Michigan. Journal of Great Lakes Research, 29, 228-241.

615

616 Bridgham, S.S., Updegraff, K., Pastor, J., 1998. Carbon, Nitrogen, and
617 Phosphorus mineralization in northern wetlands. Ecology 79, 1545-1561.

618

619 Brown, T.L., Rice, J.A., 2000. Effect of experimental parameters on the ESI
620 FT-ICR mass spectrum of fulvic acid. Analytical Chemistry 72, 384-390.

621

622 Cotner, J.B., Biddanda, B.A., Makino, W., Stets, E., 2004. Organic carbon
623 biogeochemistry of Lake Superior. Aquatic Ecosystem Health and
624 Management 7, 451-464.
625
626 D'Andrilli, J., Dittmar, T., Koch, B.P., Purcell, J.M., Marshall, A.G., Cooper,
627 W.T., 2010. Comprehensive characterization of marine dissolved organic
628 matter by Fourier transform ion cyclotron resonance mass spectrometry with
629 electrospray and atmospheric pressure photoionization. Rapid
630 Communications in Mass Spectrometry 24, 643-650.
631
632 Detenbeck, N.E., Elonen, C.M., Taylor, D.L., Anderson, L.E., Jicha, T.M.,
633 Batterman, S.L., 2004. Regional, landscape, and scale effects on Lake
634 Superior tributary water quality. Journal of the American Water Resources
635 Association. 40, 705-720.
636
637 Dittmar, T., 2008. The molecular level determination of black carbon in
638 marine dissolved organic matter. Org. Geochem. 39, 396-407.
639
640 Dittmar, T., Koch, B., Hertkorn, N, Kattner, G., 2008. A simple and efficient
641 method for the solid phase extraction of dissolved organic matter (SPE-DOM)
642 from seawater. Limnology and Oceanography: Methods 6, 230-235.
643

644 Guildford, S.J., Hecky, R.E., 2000. Total nitrogen, total phosphorus, and
645 nutrient limitation in lakes and oceans: Is there a common relationship?
646 Limnology and Oceanography 45, 1213-1223.
647
648
649 Herdendorf, C.E., 1982. Large lakes of the world. Journal of Great Lakes
650 Research 8, 379-412.
651
652 Hertkorn, N., Benner, R., Frommberger, M., Schmitt-Koplin, P., Witt, M.,
653 Kaiser, K., Kettrup, A. and Hedges, J.I., 2006. Characterization of a major
654 refractory component of marine dissolved organic matter. Geochimica et
655 Cosmochimica Acta 70, 2990-3010.
656
657 Hertkorn, N., Frommberger, M., Witt, M., Koch, B. P., Schmitt-Kopplin Ph.,
658 Perdue, E.M., 2008. Natural Organic Matter and the Event Horizon of Mass
659 Spectrometry. Analytical Chemistry 80, 8908-8919.
660
661 Hockaday, W.C., Purcell, J.M., Marshall, A.G., Baldock, J.A., Hatcher, P.G.,
662 2009. Electrospray and photoionization mass spectrometry for the
663 characterization of organic matter in natural waters: a qualitative
664 assessment. Limonology and Oceanography: Methods 7, 81-95.
665

666

667 Kido Soule, M.C., Longnecker, K., Giovannoni, S.J., Kujawinski, E.B., 2010.
668 Impact of instrument and experiment parameters on reproducibility of
669 ultrahigh resolution ESE FT-ICR mass spectra of natural organic matter.
670 Organic Geochemistry 41, 725-733.
671
672 Kim, S., Kramer, R.W., Hatcher, P.G., 2003. Graphical Method for Analysis of
673 Ultrahigh-Resolution Broadband Mass Spectra of Natural Organic Matter,
674 the Van Krevelen Diagram. Analytical Chemistry 75, 5336-5344.
675
676 Kim, S., Kaplan, L.A., Benner, R., Hatcher, P., 2004. Hydrogen-deficient
677 molecules in natural riverine water samples—Evidence for the existence of
678 black carbon in DOM. Marine Chemistry 92, 225-234.
679
680 Kim, S., Kaplan, L.A., and Hatcher, P.G., 2006. Biodegradable dissolved
681 organic matter in a temperate and a tropical stream determined from ultra-
682 high resolution mass spectrometry. Limnology and Oceanography 51, 1054-
683 1063.
684
685 Koch, B.P., Witt, M., Engbrodt, R., Dittmar, T., Kattner, G., 2005. Molecular
686 formulae of marine and terrigenous dissolved organic matte detected by

687 electrospray ionization Fourier transform ion cyclotron resonance mass
688 spectrometry. Geochimica et Cosmochimica Acta 69, 3299-3308.
689
690 Koprivnjak, J.-F., Pfrom, P.H., Ingall, E., Vetter, T.A., Schmitt-Kopplin, P.,
691 Hertkorn, N., Frommberger, M., Knicker, H., Perdue, E.M., 2009. Chemical
692 and spectroscopic characterization of marine dissolved organic matter
693 isolated using coupled reverse-osmosis-electrodialysis. Geochimica et
694 Cosmochimica Acta 73, 4215-4231.
695
696 Kruger, B.R., Dalzell, B.J., Minor, E.C., 2011. Effect of organic matter type
697 and salinity on dissolved organic matter isolation via ultrafiltration and solid
698 phase extraction. Aquatic Sciences, DOI: 10.1007/s00027-011-0189-4
699
700 Kujawinski, E.B., Behn, M.D., 2006. Automated Analysis of Electrospray
701 Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectra of
702 Natural Organic Matter. Analytical Chemistry 78, 4363-4373.
703
704 Kujawinski, E.B., Freitas, M.A., Zang, X., Hatcher, P.G., Green-Church, K.B.,
705 Jones, R.B., 2002. The application of electrospray ionization mass
706 spectrometry (ESI MS) to the structural characterization of natural organic
707 matter. Organic Geochemistry 33, 171-180.
708

Kujawinski, E.B., Del Vecchio, R., Blough, N.V., Klein, G.C., Marshall, A.G., 2004. Probing molecular-level transformations of dissolved organic matter: Insights on photochemical degradation and protozoan modification of DOM from electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. Marine Chemistry 92, 23-37.

Kujawinski, E.B., Longnecker, K., Blough, N.V., Del Vecchio, R., Finlay, L., Kitner, J.B., Giovannoni, S.J., 2009. Identification of possible source markers in marine dissolved organic matter using ultrahigh resolution mass spectrometry. Geochimica et Cosmochimica Acta 73, 4384-4399.

Lakesuperiorstreams. 2009. LakeSuperiorStreams: Community Partnerships For Understanding Water Quality and Stormwater Impacts at the Head of the Great Lakes (http://lakesuperiorstreams.org). University of Minnesota-Duluth, Duluth, MN 55812.

McIntyre, C., McRae, C., 2005. Proposed guidelines for sample preparation and ESI-MS analysis of humic substances to avoid self-esterification. Organic Geochemistry 36, 543-553.

Meyers, P.A., 1994. Preservation of elemental and isotopic source identification of sedimentary organic matter. Chem Geol. 114, 289-302.

Minor, E., Stephens, B., 2008. Dissolved organic matter characteristics within the Lake Superior watershed. Organic Geochemistry 39, 1489-1501.

Mopper, K., Zhou, X.L., Kieber, R.J., Kieber, D.J., Sikorski, R.J,, Jones, R.D., 1991. Photochemical degradation of dissolved organic carbon and its impact on the oceanic carbon cycle. Nature 353, 60-62.

Ostrom, N.E., Long, D.T., Bell, E.M., Beals, T., 1998. The origin and cycling of particulate and sedimentary organic matter and nitrate in Lake Superior. Chemical Geology 152, 13-28.

Pomeroy, L.R., 1979. Microbial roles in aquatic food webs. In: Collwell, R.R. Foster, J., Ahearn, H.L.(Ed.), Aquatic Microbial Ecology, University of Maryland, College Park, MD, pp. 85-109.

752 Reemtsma, T., 2010. The carbon versus mass diagram to visualize and exploit
753 FTICR-MS data of natural organic matter. Journal of Mass Spectrometry 45,
754 382-390.
755
756
757 Reemtsma, T., These, A., Linscheid, M., Leenheer, J., Spitzy, A., 2008.
758 Molecular and Structural Characterization of Dissolved Organic Matter from
759 the Deep Ocean by FTICR-MS, Including Hydrophilic Nitrogenous Organic
760 Molecules. Environmental Science and Technology 42, 1430-1437.
761
762
763
764 Schmidt, F., Elvert, M., Koch, B.P., Witt, M., Hinrichs, K-U., 2009. Molecular
765 characterization of dissolved organic matter in pore water of continental shelf
766 sediments. Geochimica et Cosmochimica Acta 73, 3337-3358.
767
768 Scully, N.M. and Lean, D.R.S., 1994. The attenuation of ultraviolet radiation
769 in temperate lakes. Archiv für Hydrobiologie Beiheft Ergebnisse der
770 Limnologie 43, 135-144.
771

772 Shiller, A.M., Duan, S., Van Erp, P., Bianchi, T.S., 2006. Photo-oxidation of
773 dissolved organic matter in river water and its effect on trace element
774 speciation. Limnology and Oceanography 51, 1716-1728.
775
776 Simjouw, J.P., Minor, E.C., Mopper, K., 2005. Isolation and characterization
777 of estuarine dissolved organic matter: Comparison of ultrafiltration and C-18
778 solid-phase extraction techniques. Marine Chemistry 96, 219-235.
779
780 Sleighter, R.L., Hatcher, P.G., 2007. The application of electrospray
781 ionization coupled to ultrahigh resolution mass spectrometry for the
782 molecular characterization of natural organic matter. Journal of Mass
783 Spectrometry 42, 559-574.
784
785 Sleighter, R.L., Hatcher, P.G., 2008. Molecular characterization of dissolved
786 organic matter (DOM) along a river to ocean transect of the lower
787 Chesapeake Bay by ultrahigh resolution electrospray ionization Fourier
788 transform ion cyclotron resonance mass spectrometry. Marine Chemistry
789 110, 140-152.
790
791 Sleighter, R.L., McKee, G.A., Liu, Z., Hatcher, P.G., 2008. Naturally present
792 fatty acids as internal calibrants for Fourier transform mass spectra of
793 dissolved organic matter. Limnology and Oceanography: Methods 6, 246-253.

794

795 Stenson, A.C., Landing, W.M., Marshall, A.G., and Cooper, W.T., 2002.
796 Ionization and fragmentation of humic substances in electrospray ionization
797 Fourier transform ion cyclotron resonance mass spectrometry. Analytical
798 Chemistry 74, 4397-4409.

799

800 Sterner, R.W., Smutka, T.M., McKay, R. Michael. L., Xiaoming, Q., Brown,
801 E.T., Sherrell, R.M., 2004. Phosphorus and trace metal limitation of algae
802 and bacteria in Lake Superior. Limnology and Oceanography 49, 495-507.

803

804 Sterner, R. W., Anagnostou, E., Brovold, S., Bullerjahn, G.S., Finlay, J.C.,
805 Kumar, S., McKay, R.M.L. and Sherrell R.M., 2007. Increasing stoichiometric
806 imbalance in North America's largest lake: Nitrification in Lake Superior.
807 Geophysical Research Letters 34, L10406, doi: 10.1029/2006GL028861.

808

809 Urban, N.R., Auer, M.T., Green, S.A., Lu, X., Apul, D.S., Powell, K.D., Bub,
810 L., 2005. Carbon cycling in Lake Superior. Journal of Geophysical Research
811 110, C06S90.

812

813

814 **Table 1**
815 Sampling details and recovery from C_{18} extracts as proportion of initial sample
816 colored DOM (CDOM) and initial sample DOC content.
817

Sample	Latitude & Longitude	Description	Date Collected	C18 recovery (% CDOM)	C18 recovery (% DOC)
Toivola Swamp	47º20'N 93º38'W	Stream in swamp	3 May 2007	43	43
St. Louis River	46º37'N 92º17'W	4th order brown river, sampled upstream of most seiche impacts	14 May 2007	36	48
Canal Park	46º47'N 92º05'W	Duluth Harbor (upstream of canal entrance)	21 May 2007	43	51
Lake Superior	47º09'N 91º17'W	Offshore Lake Superior sample. Sample depth 50 m, total water depth >300m	6 June 2007	50	21
Amity Creek Low Flow	46º50'N 92º01'W	Amity Creek Low Flow (flow at 0.055 m^3s^{-1})*. North Shore forested stream	8 May 2007	80	N.D.
Amity Creek Higher Flow	46º50'N 92º01'W	Amity Creek Higher Flow (flow at 0.210 m^3s^{-1})[a], North Shore	27 May 2007	75	N.D.

		forested stream			
Brule River	46°33'N, 91°35'W	South Shore forested river	20 June 2008	37	25

[a]Stream gage data (Lake Superior Streams, 2009) was taken at a site upstream of our sampling location.

Table 2

Proportion of peaks shared between replicate injections. SD stands for sample standard deviation.

Sample site	Treatment	S/N=3 Mean # Peaks (SD)	S/N=3 Peaks shared (%)	S/N=5 Mean # Peaks (SD)	S/N=5 Peaks shared (%)
Canal Park	Triplicate	3393 (745)	65	2450 (222)	87
Amity Creek (Low Flow)	Triplicate	3462 (407)	64	2696 (106)	92
Suwannee River Fulvic Acid (IHSS)	March 2009, March 2010	3613 (83)	62	2551 (97)	91
All Superior-watershed sites	Single	3035 (619)	22	2637 (323)	60

Table 3

Elemental data from formula assignments. Average values and their standard deviation are reported for samples measured in triplicate. The subscript $_w$ indicates magnitude-averaged values; the subscript $_n$ indicates number-averaged values. All values not indicated with a subscript were derived from presence-absence data.

	Lake Superior	Amity Creek low flow Average (SD)	Brule River	St. Louis River	Canal Park Average (SD)	Toivola Swamp	Amity Creek higher flow
Total number of *m/z* values	4220	3461 (407)	4008	3015	2765 (449)	3283	2348
%Formulae	98.4	99.3 (0.3)	96.8	99.4	99.4 (0.5)	98.4	99.1
DBE	13.1	12.48 (0.08)	10.8	12.6	11.6 (0.2)	14.2	12.1
DBE/C	0.54	0.51 (0.00)	0.52	0.51	0.499 (0.001)	0.59	0.53
H/C_w	1.22	1.22 (0.05)	1.28	1.20	1.23 (0.00)	1.08	1.18
O/C_w	0.377	0.41 (0.02)	0.416	0.396	0.402 (0.002)	0.386	0.406
N/C_w	0.13	0.10 (0.01)	0.10	0.08	0.08 (0.01)	0.11	0.09
P/C_w	0.008	0.002 (0.000)	0.004	0.001	0.002 (0.000)	0.005	0.001
S/C_w	0.070	0.001 (0.000)	0.001	0.001	0.001 (0.000)	0.001	0.000
H/C_n	1.17	1.22 (0.07)	1.20	1.16	1.19 (0.00)	1.04	1.15
O/C_n	0.36	0.43 (0.03)	0.41	0.40	0.41 (0.00)	0.39	0.41
N/C_n	0.15	0.11 (0.01)	0.13	0.10	0.10 (0.01)	0.13	0.10
P/C_n	0.01	0.00 (0.00)	0.01	0.00	0.00 (0.00)	0.01	0.00
S/C_n	0.00	0.00 (0.00)	0.00	0.00	0.00 (0.00)	0.00	0.00
F_N (%)	41	30 (3)	32	22	22 (3)	33	22
F_P (%)	11.3	6.0 (0.4)	12.0	4.4	4. (1)	9.5	3.8
F_S (%)	3.4	2.3 (0.4)	2.8	1.5	1.8 (0.3)	2.6	1.5
F_{13C} (%)	20	20.3 (0.2)	19	21	20 (1)	12	18
% CHO	56	69 (3)	65	77	77 (3)	64	77
% CHON	29	23 (2)	18	17	16 (2)	23	17
% CHONP	8.9	4.6 (0.2)	9.4	3.3	3 (1)	8.2	2.9
% CHOS	1.2	0.58 (0.09)	0.70	0.13	0.4 (0.1)	1.1	0.34
%CHONSP	0.85	0.98 (0.04)	0.82	0.96	0.97 (0.03)	0.85	0.81
% CHOP	0.45	0.09 (0.03)	0.35	0.03	0.1 (0.1)	0.12	0.09
% CHONS	0.50	0.4 (0.2)	0.67	0.27	0.2 (0.2)	0.12	0.26

857

858 **Fig. 1.** Station locations including Toivola Swamp, St. Louis River, Canal
859 Park, Amity Creek, Lake Superior and Brule River, along with land use in
860 the watershed.

861

862 **Fig. 2.** Cluster diagrams for presence-absence data from all negative ion
863 mode spectra processed with A. a signal-to-noise ratio of >3 or B. a signal to
864 noise ratio of >5. For both A and B, the *X*-axis displays dissimilarity of
865 samples.

866

867 **Fig. 3.** Negative ion mode spectrum of DOM from Canal Park.

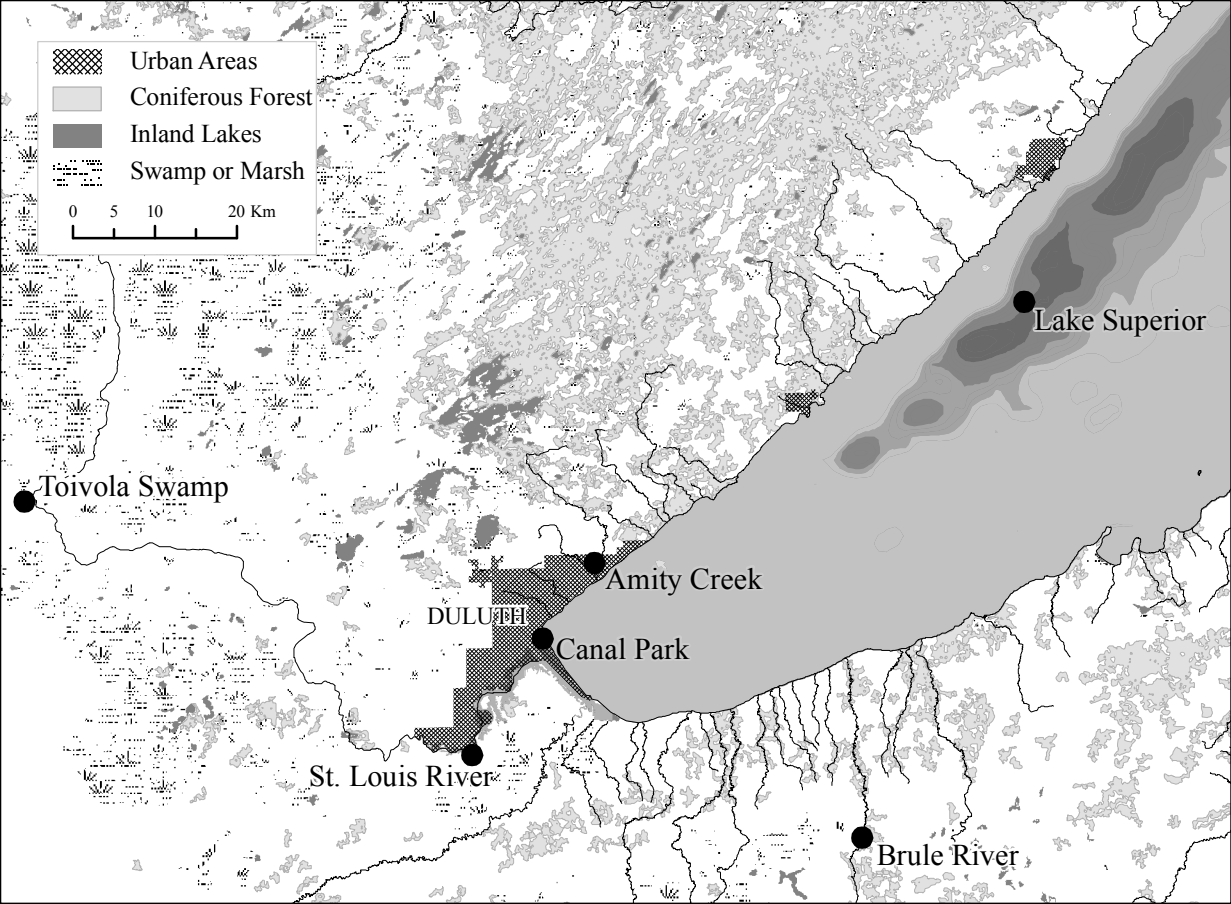
868

869 **Fig. 4.** Van Krevelen diagrams based upon presence/absence data from
870 negative ion mode ESI FT-ICR-MS data at S/N>3. Ellipses indicate
871 elemental ratios consistent with compound classes listed (Kujawinski et al.,
872 2009; Kujawinski and Behn, 2006; Sleighter and Hatcher, 2007; Hockaday et
873 al., 2009), but do not imply absolute identifications of these compound
874 classes.

875

876 **Fig. 5.** Van Krevelen diagram of compound assignments unique to each site
877 (based upon presence/absence data from negative ion mode ESI FT-ICR-MS
878 at S/N>3).

879


Urban Areas
Coniferous Forest
Inland Lakes
Swamp or Marsh
0 5 10 20 Km
Lake Superior
Toivola Swamp
Amity Creek
DULUTH
Canal Park
St. Louis River
Brule River

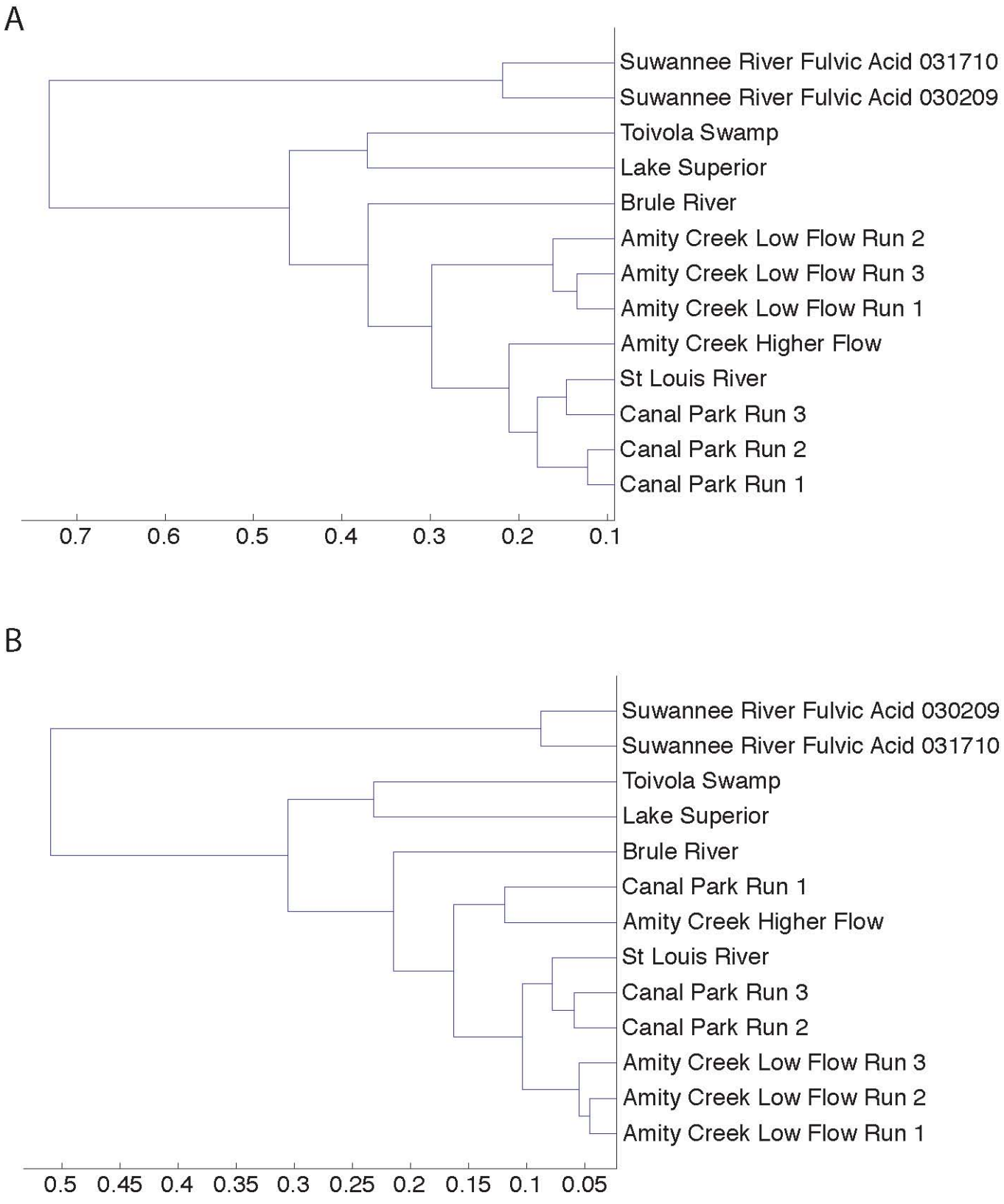
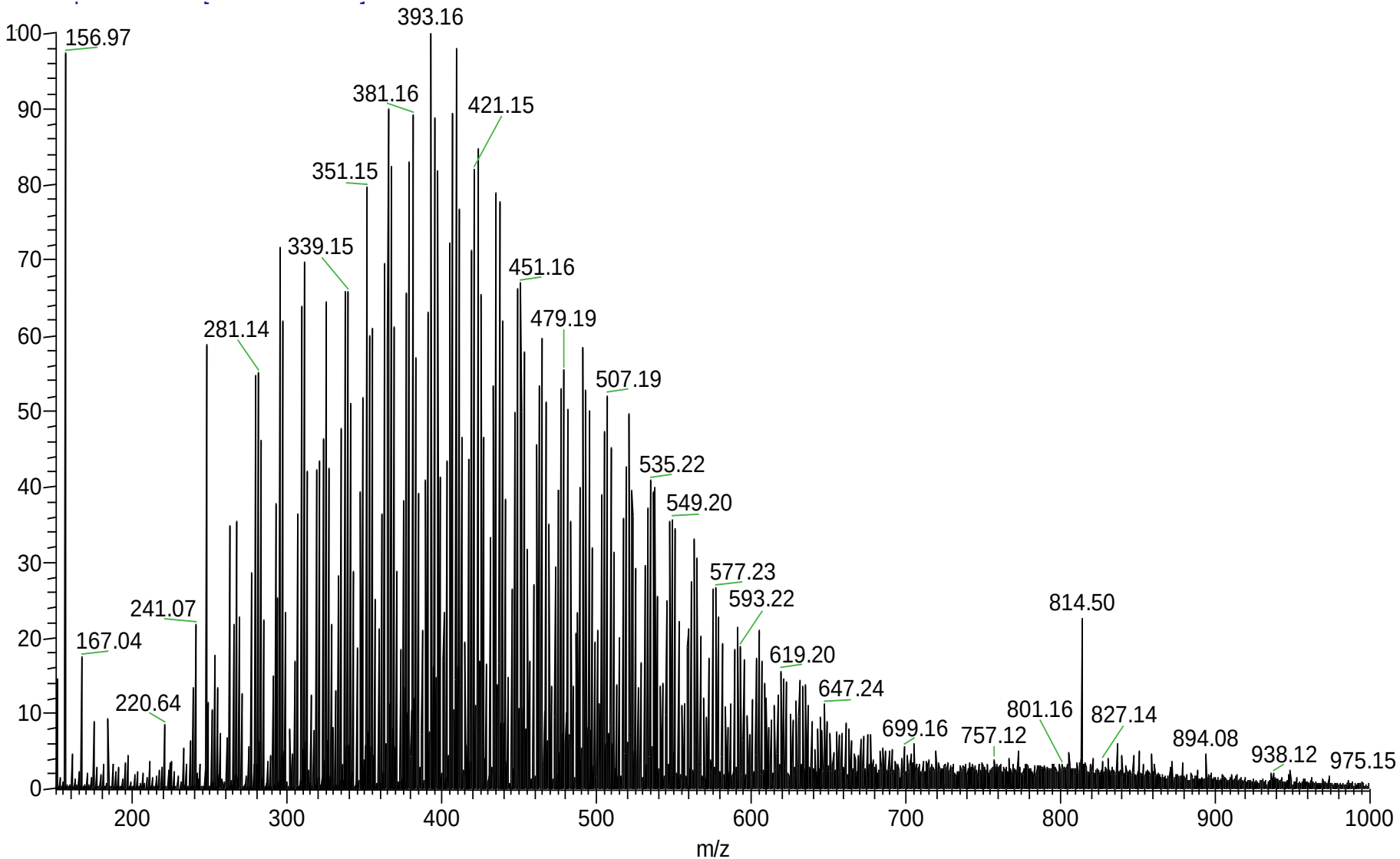
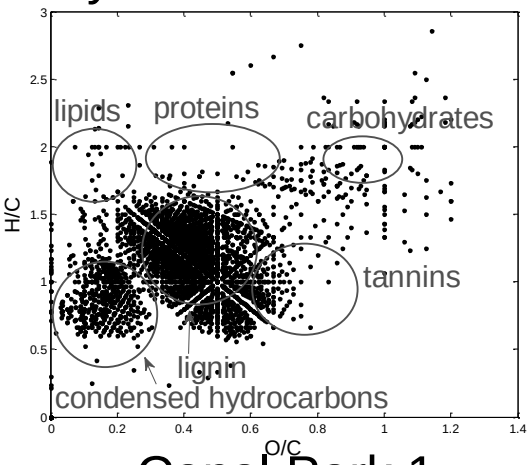

A
Suwannee River Fulvic Acid 031710
Suwannee River Fulvic Acid 030209
Toivola Swamp
Lake Superior
Brule River
Amity Creek Low Flow Run 2
Amity Creek Low Flow Run 3
Amity Creek Low Flow Run 1
Amity Creek Higher Flow
St Louis River
Canal Park Run 3
Canal Park Run 2
Canal Park Run 1
0.7 0.6 0.5 0.4 0.3 0.2 0.1
B
Suwannee River Fulvic Acid 030209
Suwannee River Fulvic Acid 031710
Toivola Swamp
Lake Superior
Brule River
Canal Park Run 1
Amity Creek Higher Flow
St Louis River
Canal Park Run 3
Canal Park Run 2
Amity Creek Low Flow Run 3
Amity Creek Low Flow Run 2
Amity Creek Low Flow Run 1
0.5 0.45 0.4 0.35 0.3 0.25 0.2 0.15 0.1 0.05

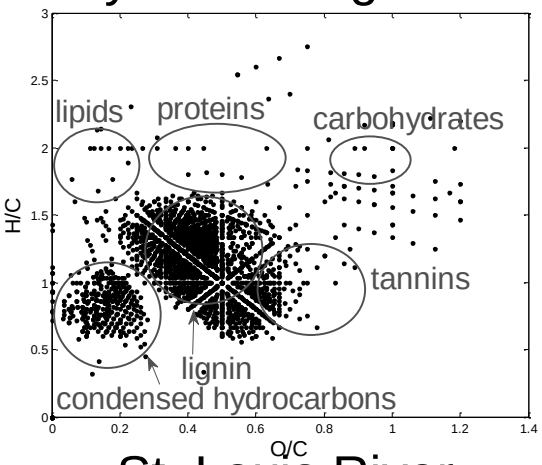
Figure 2


100
90
80
70
60
50
40
30
20
10
0
200
300
400
500
600
700
800
900
1000
m/z
156.97
393.16
381.16
421.15
351.15
339.15
451.16
281.14
479.19
507.19
535.22
549.20
577.23
593.22
241.07
167.04
619.20
220.64
647.24
699.16
757.12
801.16
814.50
827.14
894.08
938.12
975.15

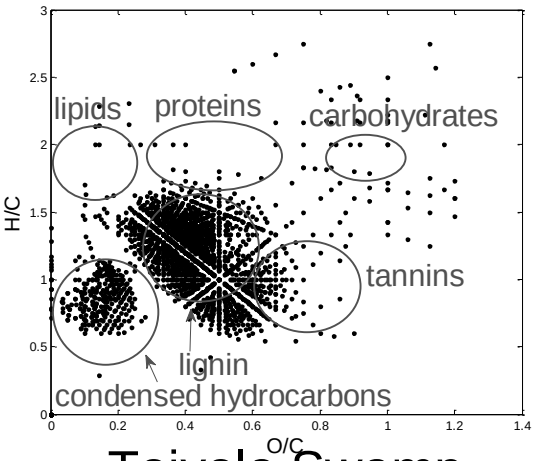
Amity Creek Lower Flow 1



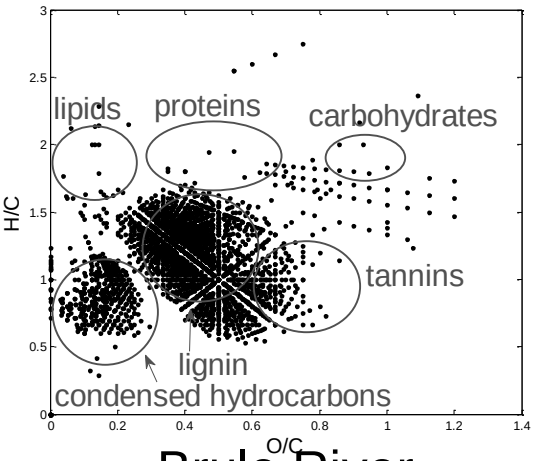
Amity Creek Higher Flow



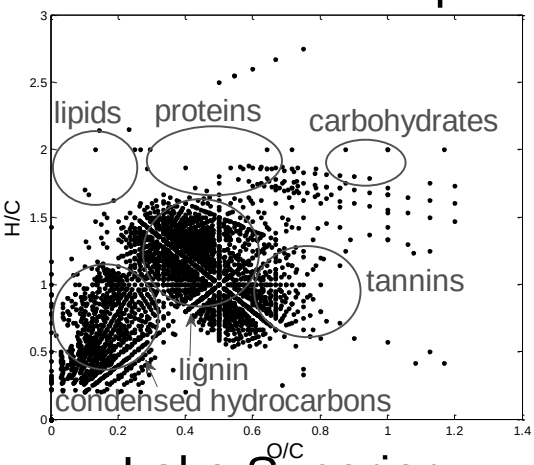
Canal Park 1



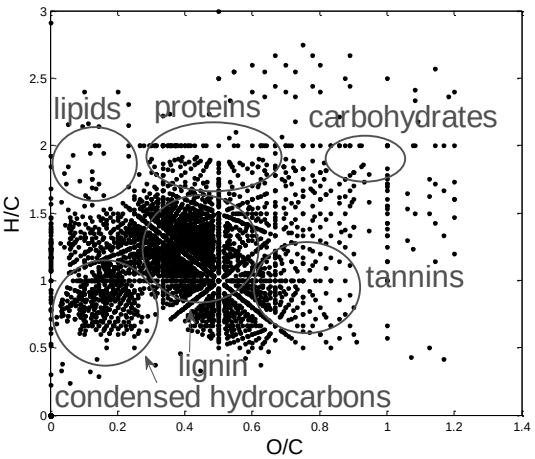
St. Louis River



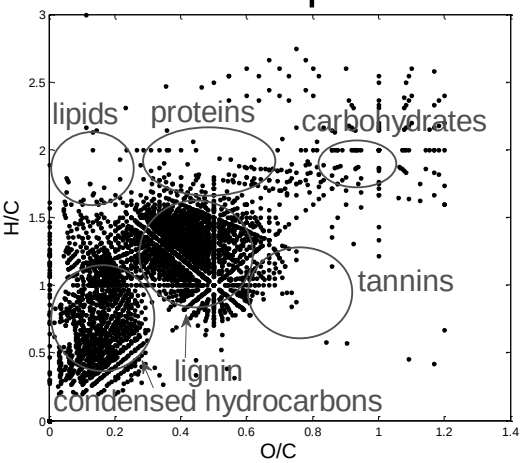
Toivola Swamp



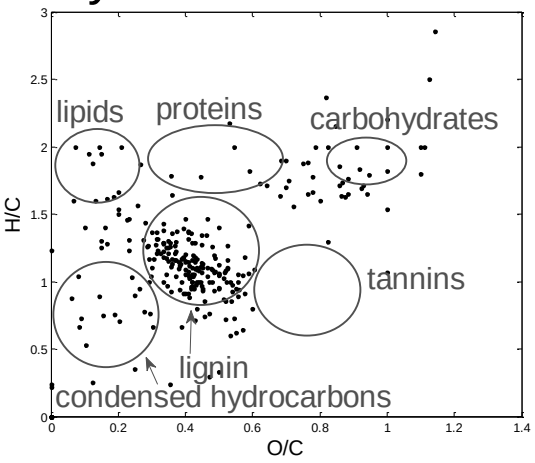
Brule River



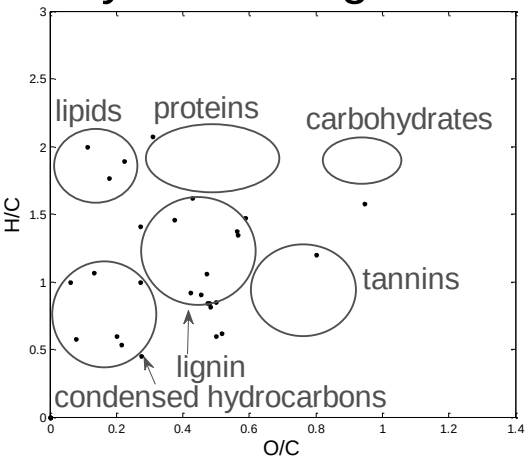
Lake Superior



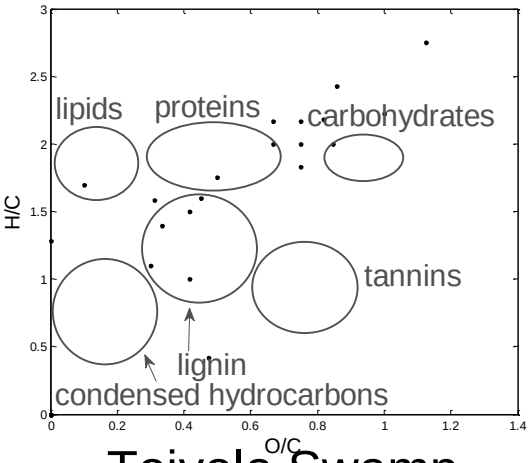
Amity Creek Lower Flow 1



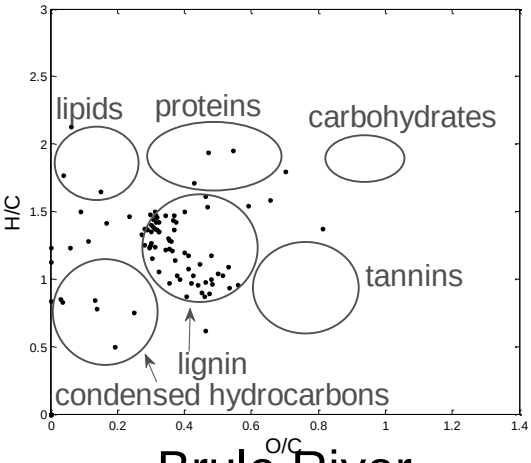
Amity Creek Higher Flow



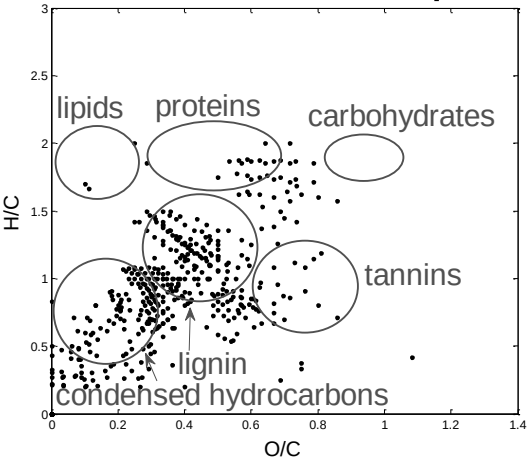
Canal Park 1



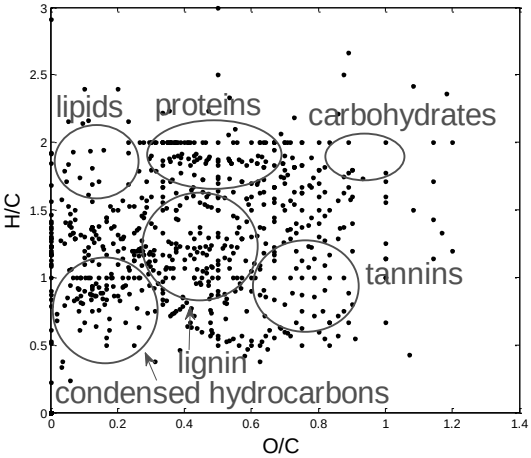
St. Louis River



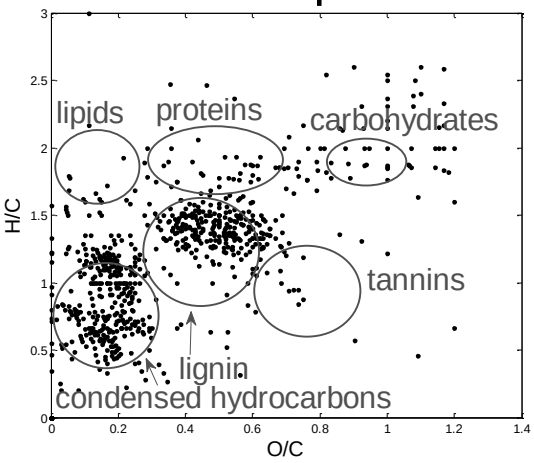
Toivola Swamp



Brule River



Lake Superior



Peaks and formula assignments used to calibrate peaks in MIDAS.

m/z Uncalibrated Molecular Ion	*m/z* Calibrated Molecular Ion	Calculated Molecular ion	C	H	O	Average Intensity
279.050986429	279.051028	279.050477706	13	12	7	7
321.097946429	321.097975	321.097427898	16	18	7	7
363.144913571	363.144934	363.144378090	19	24	7	50
405.082724286	405.082733	405.082171764	19	18	10	10
445.114028571	445.114011	445.113471892	22	22	10	16
487.160982857	487.160951	487.160422084	25	28	10	21
*541.156263077	541.156265	541.155730636	24	30	14	9
583.203237692	583.203244	583.202680828	27	36	14	12
625.250173333	625.250191	625.249631020	30	42	14	9

* After calibrating the peaks using MIDAS, the formula assignment for the peak at 541.16 was reassigned from $C_{24}H_{30}O_{14}$, to $C_{22}H_{18}O_4N_{14}$. The difference between the calculated molecular ions for these two formulae is 0.0000106.