Ultrahigh Resolution Mass Spectrometry and Indicator Species Analysis to Identify Marker Components of Soil- and Plant Biomass-Derived Organic Matter Fractions

ARTICLE *in* ENVIRONMENTAL SCIENCE & TECHNOLOGY · OCTOBER 2010

Impact Factor: 5.33 · DOI: 10.1021/es101089t · Source: PubMed

5 AUTHORS, INCLUDING:

Tsutomu Ohno
University of Maine
61 PUBLICATIONS **1,889** CITATIONS

SEE PROFILE

Ultrahigh Resolution Mass Spectrometry and Indicator Species Analysis to Identify Marker Components of Soil- and Plant Biomass-Derived Organic Matter Fractions

TSUTOMU OHNO,*,† ZHONGQI HE,‡ RACHEL L. SLEIGHTER,§ C. WAYNE HONEYCUTT,‡ AND PATRICK G. HATCHER§

Department of Plant, Soil, and Environmental Sciences, 5722 Deering Hall, University of Maine, Orono, Maine 04469-5722, United States, New England Plant, Soil, and Water Laboratory, USDA-ARS, Orono, Maine 04469, United States, and Department of Chemistry and Biochemistry, Old Dominion University, Norfolk, Virginia 23529, United States

Received April 6, 2010. Revised manuscript received September 20, 2010. Accepted October 1, 2010.

The chemical properties of organic matter affect important soil processes such as speciation, solubilization, and transport of plant nutrients and metals. This work uses ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry to determine the molecular composition of three organic matter fractions of soils and aqueous extracts of crop biomass. Comparison of the van Krevelen plots allowed tracking the changes in organic matter with increasing humification. Aqueous plant biomass extracts contain a diverse mixture of lipids, proteins, and lignins. Soil aqueous extracts were marked by increases in lignin and carbohydrate components and decrease in the protein component as compared to the plant extract. Refractory humic acid fractions were marked by decrease in the lignin component and increases in the lipid and condensed aromatic components. The multivariate indicator species analysis was used to identify marker components of the four organic matter types investigated. The plant extract group had 772 marker components compared to 237 for soil aqueous extract, 92 for mobile humic acid, and 418 for calcium humic acid. This study demonstrates that ultrahigh resolution mass spectrometry and multivariate methods can be used to identify marker components to gain a molecular-scale description and understanding of C dynamics.

Introduction

The characterization of soil organic matter (SOM) is a fundamental challenge given its diverse chemical complexity and its importance in many vital soil ecosystem processes. The chemical properties of SOM affect the speciation, bioavailability, transport, and distribution of both plant nutrients and trace elements (*1–3*). SOM is primarily formed through a humification process involving the conversion of plant-derived biomolecules to a more condensed chemical structure by microbial synthesis (*4–6*). The formation of SOM is influenced by vegetation, macro- and micro-organisms, climate, and soil physical properties (*7*). SOM contents are generally higher with increasing mean annual precipitation and decreasing mean annual temperature (*8*). Soil texture also affects SOM content, with the silt and clay fractions providing protection from microbial attack through adsorption and aggregation mechanisms (*9*).

Although SOM exists on a humification continuum, operationally defined extractions have been extensively used to fractionate SOM into different classes to reduce heterogeneity and elucidate structure and function of these fractions. Water-extractable organic matter (WEOM) represents the most labile fraction of SOM and approximates the dissolved organic matter found in soil solution (*10*). SOM has traditionally been extracted using a strong base to isolate humic substances (*11*). A recent extraction protocol uses NaOH and binding of SOM to polyvalent soil cations to fractionate SOM into an unbound, mobile humic acid (MHA) of intermediate lability and humification, as well as a more refractory and humified cationic-bound calcium humate (CaHA) fraction (*12*).

Whereas a multitude of chemical characterization techniques have been employed to describe SOM fractions, a relatively new technique called ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry (ESI-FT-ICR-MS) has emerged as a powerful state-of-the-art method for molecular characterization of environmentally significant molecules (*13*). The superior performance factors of ESI-FT-ICR-MS make molecular determinations possible, with the capability of resolving up to dozens of compounds per nominal mass, and it is ideal for molecular-level examination of humic substances (*13–16*). The interpretation of ultrahigh resolution ESI-FT-ICR-MS data benefit from the use of graphical analysis, to simplify and visually display the data in a chemically significant fashion. The van Krevelen diagram is one such approach where the H/C and O/C ratio of each component is plotted, which allows categorization of the components into general biomolecular compound classes (*17*). Kujawinski and co-workers (*18*) recently applied the Indicator Species Analysis (ISA) statistical method, frequently used in the ecology discipline, to identify chemical species indicative of groups of samples obtained from a transect of the North Atlantic Ocean. ISA calculates an Indicator Value (IV) for each species (i.e., each peak in the mass spectrum) as the product of its relative abundance and relative frequency (*19*). Combining the ISA results with molecular formula calculation and visualization transformations, such as van Krevelen diagrams, should assist in describing the operational fractions of SOM to a molecular scale previously unattainable.

Greater understanding of the chemical nature of SOM will help in establishing judicious management of these soil resources, which potentially may lead to enhanced environmental quality and function. The specific objective of this study is to characterize the chemical composition of the WEOM, MHA, and CaHA fractions of ten soils collected from a variety of soil types and climate regions of the United States

* Corresponding author phone: 207-581-2975; e-mail: ohno@maine.edu.
† University of Maine.
‡ USDA-ARS.
§ Old Dominion University.

10.1021/es101089t

and aqueous extracts of seven crop species that are representative of the stock organic inputs for the SOM formation process.

Materials and Methods

Soil and Plant Sampling. Field soil samples were collected from USDA-ARS research centers across the United States. All soil samples were taken in areas that had received no N inputs for several years and no manure additions for at least 10 years. Soils at each location were randomly sampled by collecting 12 cores using a sliding drop hammer. The field-moist soil was placed in a cooler containing frozen ice packs and shipped to Maine. The Catlin soil sample from Illinois was an exception, with the soil collected to an approximate 15-cm depth using a shovel and storage at ambient temperature for one week during transport. Plant shoot biomass was obtained from field-grown alfalfa (*Medicago sativa* L.), corn (*Zea mays* L.), crimson clover (*Trifolium incarnatum* L.), hairy vetch (*Vicia villosa* L.), lupin *(Lupinus albus* L.), soybean (*Glycine max* L. Merr.), and wheat (*Triticum aestivum* L.), air-dried, and ground to pass through a 1-mm sieve. The alfalfa, crimson clover, hairy vetch, and lupin were field grown and harvested at full flowering stage. The corn, soybean, and wheat were sampled as crop residue left after harvest. Details about the organic matter extractions (including reproducibility of the extractions) and fluorescence EEM/PARAFAC analysis are given in the Supporting Information.

ESI-FT-ICR-MS Analysis. All ESI-FT-ICR-MS analyses were conducted with the 12 T Bruker Daltonics Apex Qe FT-ICR-MS housed at the College of Sciences Major Instrumentation Cluster at Old Dominion University. The samples were diluted with LC-MS grade methanol to give a final sample composition of 50:50 (v/v) methanol/water. To increase the ionization efficiency, ammonium hydroxide was added immediately prior to ESI to bring the pH up to 8. Samples were introduced by a syringe pump providing an infusion rate of 120 μL hr^{-1} and analyzed in negative ion mode with electrospray voltages optimized for each sample. Previous studies have shown that negative ion mode avoids the complications of the positive ion mode in which alkali metal adducts, mainly Na^+, are observed along with protonated ions (*20, 21*). Ions (in the range of 200−2000 m/z) were accumulated in a hexapole for 1.0 s before being transferred to the ICR cell. Exactly 300 transients, collected with a 4 MWord time domain, were added, giving about a 30 min total run time. The summed free induction decay (FID) signal was zero-filled once and Sine-Bell apodized prior to fast Fourier transformation and magnitude calculation using the Bruker Daltonics Data Analysis software. Prior to data analysis, all samples were externally calibrated with a polyethylene glycol standard and internally calibrated with naturally present fatty acids within the sample. A molecular formula calculator developed at the National High Magnetic Field Laboratory in Tallahassee, FL (Molecular Formula Calc v.1.0 NHMFL, 1998) was used to generate empirical formula matches for the resolved peaks using C (8−50 atoms), H (8−100 atoms), O (1−30 atoms), N (0−5 atoms), and S (0−2 atoms) as the limiting atomic values. Only m/z values with a signal-to-noise above 4 were used in the molecular formula calculator. The resulting formula list was passed through a MATLAB routine to constrain formulas to chemically feasible formulas. For species with $m/z > 400$, two or three formula assignments may be listed for a particular m/z value based on a preselected error window (1 ppm). In these cases, one formula assignment was selected based first on the lowest number of N and S atoms and the lowest deviation between observed and calculated m/z values (*18*).

Indicator Species Analysis. The sample set consisted of the CaHA, MHA, and WEOM fractions from each of the ten soils and the plant WEOM from the plant samples. These four organic matter (OM) groupings were used as the ISA required *a priori* group assignment for the organic matter type. A MATLAB routine developed by the Kujawinski group was used to generate a matrix holding all m/z values found in the 37 mass spectra and their presence (1) or absence (0) for each sample. This 37 × 12743 data matrix was exported to PC-ORD v 5.25 which was used to conduct ISA. The IV threshold of 50 and p-value less than 0.05 generated from a Monte Carlo simulation test was used to indentify indicator species for each group (*18*).

Results and Discussion

EEM/PARAFAC Analysis of the Soil Organic Matter Fractions. The excitation and emission spectral loadings of the four organic matter components identified by EEM/PARAFAC fluorescence are shown in Figure 1a−d. The loadings of components 1 and 2 are similar to those found in IHSS humic and fulvic materials (*22*). Component 3 loadings are similar to excitation and emission spectra of lignin isolated from trees (*23, 24*). Component 4 loadings are associated with tryptophan containing proteins (*25*). The relative fraction distribution of the four components is shown in Figure 1e−h. The humic-like component 1 content decreases steadily along the CaHA→MHA→Soil WEOM→Plant WEOM, indicating that the four OM groups occur along a humification gradient. Component 2 content was nearly equal in the soil and humic OM groups, which suggests that it was an intermediate transition fraction. The lignin-like component 3 decreases steadily from the CaHA→Plant WEOM group, suggesting that the lignin-like fraction of OM is slowly being altered by decomposition processes, whereas the free protein-like component 4 is rapidly consumed and is not present as a major fluorophore in the soil OM and humic material.

Ultrahigh Resolution Mass Spectra. The negative-ion ESI-FT-ICR mass spectra for the samples within an OM type were visually similar across all samples, with thousands of peaks being detected across the m/z range of 200−800 and clusters of 10−20 peaks being detected at each nominal mass. ESI(−) primarily ionizes acidic and oxygen-rich functionalities, making it the optimal selection for these analyses. Although replicate injections were not done during this study, reproducibility has been tested on other DOM samples to ensure that samples can be reliably compared to one another. Sleighter et al. (*26*) reported 6 replicate injections of 1 sample over the course of 31 days to ensure that samples analyzed using the same instrumental parameters but examined on different days could be compared. Based on hierarchal cluster analysis and principal component analysis, the 6 mass spectra were considered to be nearly identical when compared to a variety of other samples. Because instrumental parameters were held constant for the samples in this study and each sample was biased in the same fashion, the spectra and molecular formula assignments can be compared. The aqueous alfalfa extract and the three SOM fractions for the Adkins soil are shown as representative spectra in Figure 2a−d. The molecular weight (MW) range decreased as humification increased from the aqueous alfalfa extract to the Adkins CaHA fraction. This pattern was found across all plant and soil OM fractions with the average highest MW peak found at 822 for plant WEOM, 683 for soil WEOM, 701 for MHA, and 665 for CaHA. This suggests that the higher MW biomolecules are transformed into lower MW components as a result of the humification process or metabolized into microbial biomass. The ~225 to 800 MW ranges found for these four terrestrially derived OM fractions are similar to the MW ranges reported for dissolved organic matter found in marine (*27*), estuarine (*16*), and riverine (*28*) sources. The MWs when viewed as a bulk average of the components for the different OM types were similar to the MW range pattern,

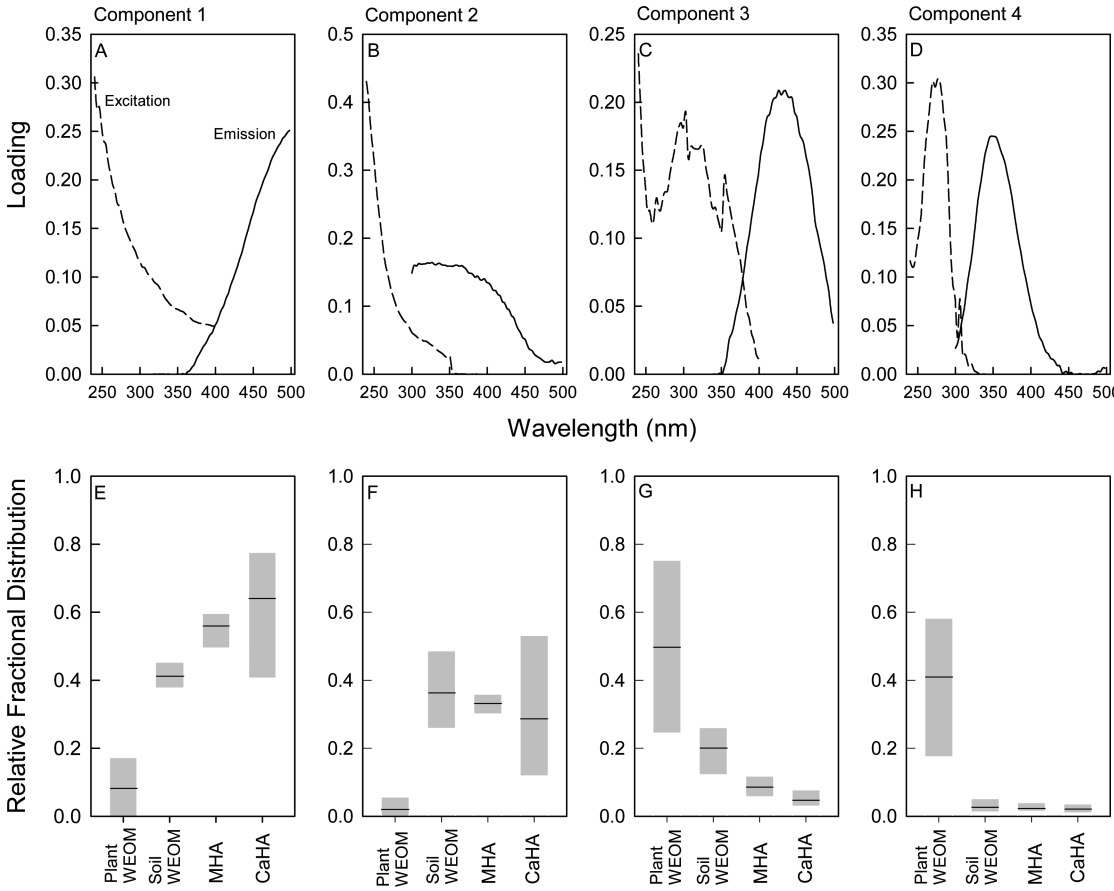

FIGURE 1. Excitation and emission spectral loadings of the non-negativity constrained four-component PARAFAC model (a−d) and the relative fractional distribution of the four components for the organic matter isolated from water-extractable organic matter from alfalfa biomass, water-extractable organic matter from Adkins soil, mobile humic acid from Adkins soil, and refractory calcium humic acid from Adkins soil (e−h).

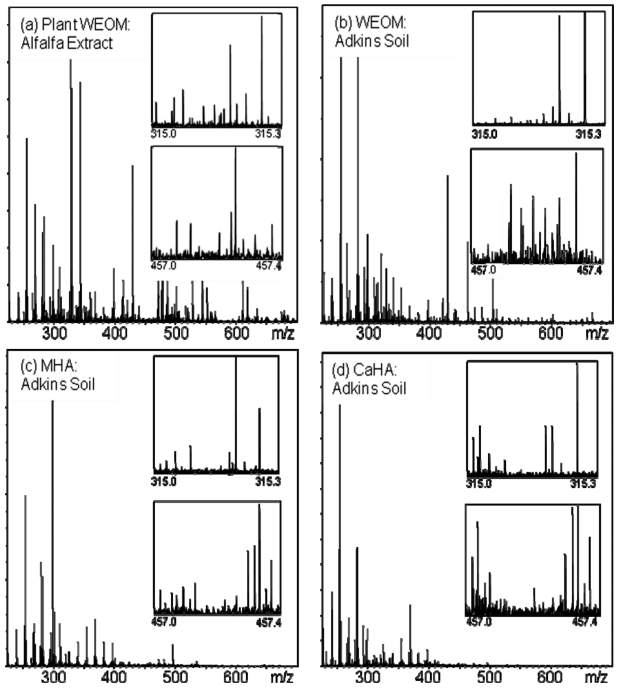

FIGURE 2. Representative ESI-FT-ICR mass spectra from (a) water-extractable organic matter from alfalfa biomass, (b) water-extractable organic matter from Adkins soil, (c) mobile humic acid from Adkins soil, and (d) refractory calcium humic acid from Adkins soil.

with the plant WEOM MW being significantly higher than the soil WEOM (Table 1). These data are not consistent with the "polymer model" of soil humification where plant biomass degradation products undergo polymerization reactions to form humic substances (*11*). The data here support the view recently proposed by Sutton and Sposito (*29*) that humic substances are associations of low MW components held together by hydrophobic interactions and hydrogen bonding.

The ability of ultrahigh resolution ESI-FT-ICR-MS to accurately determine m/z to five decimal places allows for the calculation of an elemental formula for each peak (i.e., component), based on the atomic weights of the elements assumed to be present in OM. The bulk average CHONS formula for each of the 4 different OM types is shown in Table 1. Analysis of variance revealed significant differences ($p = 0.05$) between the OM types for each of the parameters. The average number of C, which is the structural backbone of OM, was found in a narrow range between 20.0 and 23.6. The average O content, which is key to acidic functional groups of OM, was also found to be in a narrow range of 6.1−7.4 between the four OM types. The H content and the H/C ratio of plant and soil WEOM and the soil MHA fraction did not significantly differ from each other, but were significantly different from the values for the refractory CaHA fraction. The formula $((2 \times \text{no. C}) + 2 - \text{no. H} + \text{no. N})/2$ was used to calculate DBE which is the sum of double bonds and rings for a molecule (*13*). The DBE values for the two WEOM fractions were not significantly different, but increasing DBE values in the progression of soil WEOM to MHA to CaHA were significant, indicating the greater aromaticity of the organic matter with humification. Interestingly, both

TABLE 1. Bulk Organic Matter Average Formula Atomic Numbers, O/C and H/C Ratios, Molecular Weight, Double Bond Equivalents for Plant Biomass Water Extractable Organic Matter, Soil Water Extractable Organic Matter, Mobile Humic Acid, and Calcium Humic Acid Fractions (Ranges of Parameter Values Are Shown in Parentheses and the Least Significant Difference Values Were Calculated at the $p = 0.05$ Level)

OM Type	C	H	O	N	S	O/C	H/C	*m/z*	DBE
plant WEOM	22.7 (21.8−24.9)	33.3 (31.1−36.1)	7.43 (6.11−8.40)	0.79 (0.44−1.25)	0.22 (0.15−0.29)	0.36 (0.30−0.41)	1.50 (1.36−1.61)	431 (415−443)	7.38 (5.83−104)
soil WEOM	20.0 (19.0−20.6)	29.5 (24.5−35.7)	7.36 (5.44−8.39)	0.24 (0.14−0.40)	0.16 (0.10−0.24)	0.40 (0.28−0.47)	1.50 (1.30−1.74)	395 (380−417)	6.35 (3.89−7.85)
soil MHA	23.6 (20.8−23−9)	31.2 (19.6−37.1)	6.09 (4.72−7.01)	0.27 (0.14−0.57)	0.18 (0.10−0.35)	0.28 (0.23−0.34)	1.34 (0.92−1.65)	420 (381−429)	9.16 (5.42−14.5)
soil CaHA	22.2 (19.9−24.2)	23.6 (16.6−33.3)	6.50 (5.78−7.62)	0.49 (0.19−1.11)	0.31 (0.14−0.62)	0.33 (0.26−0.38)	1.12 (0.83−1.47)	410 (389−448)	11.6 (6.71−16.2)
LSD	1.2	5.0	0.72	0.18	0.08	0.04	0.18	20	2.15

N and S, which are macro-nutrients generally supplied to plants through mineralization of organic matter, increase in average content as the soil OM becomes more humified (Table 1).

van Krevelen Analysis. van Krevelen diagrams are graphical plots of the elemental H:C versus O:C ratios of the molecular formulas, which assign the components into chemical classes in the van Krevelen space (*13, 17*). The stoichiometric ranges used to establish boundaries of the classification space for the components found in natural organic matter are quite variable in the literature (*13,17,18,30,31*). The van Krevelen space for this study was divided into six discrete regions, modified from the diagram proposed by Hockaday et al. (*32*): lipids (H:C = 1.5−2.0, O:C = 0−0.3); proteins (H:C = 1.5−2.2, O:C = 0.3−0.67); lignins (H:C = 0.7−1.5, O:C = 0.1−0.67); carbohydrates (H:C = 1.5−2.4, O:C = 0.67−1.2); unsaturated hydrocarbons (H:C = 0.7−1.5, O:C = 0−0.1); and condensed aromatic structures (H:C = 0.2−0.7, O:C = 0−0.67). The van Krevelen plots derived from peaks determined in the Figure 2 mass spectra of the alfalfa- and Adkins soil-derived OM are shown in Figure 3a−d, and are separated into CHO, CHON, CHOS, and CHONS elemental formula classes. The changes in chemical composition of the OM as it undergoes humification from plant-released biomolecules to refractory humic substances are clearly shown. The aqueous alfalfa extract components show a diverse distribution pattern, with the majority of the components in the lipid, protein, and lignin classifications (Figure 3a). The soil WEOM, which represents the most labile form of SOM, has lignin components as the dominant type and decreased occurrence of lipid and protein molecules as compared to the alfalfa extract (Figure 3b). The less refractory MHA fraction has two clusters of points: one in the lipid space and the other spanning the border of the lignin and condensed aromatic space (Figure 3c). The more refractory CaHA fraction also shows two clusters: one for lipid com-

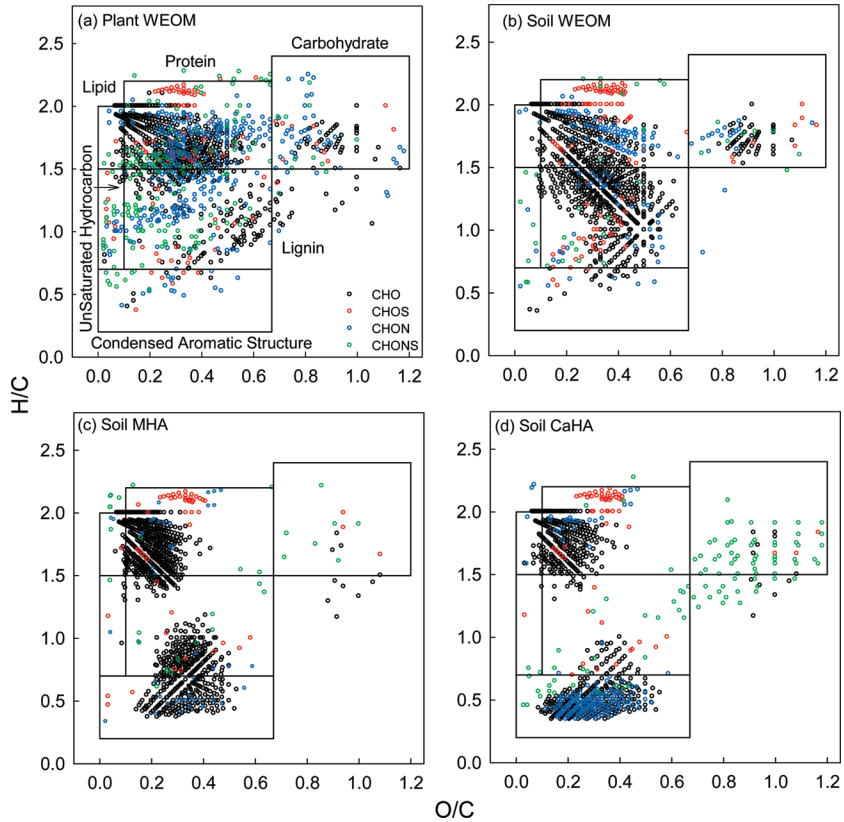

FIGURE 3. van Krevelen diagrams from the mass spectra of (a) water-extractable organic matter from alfalfa biomass, (b) water-extractable organic matter from Caribou soil, (c) mobile humic acid from Caribou soil, and (d) refractory calcium humic acid from Caribou soil. Boxes overlain on the plots indicate major biomolecular compound classes: lipids, proteins, carbohydrates, unsaturated hydrocarbons, lignins, and condensed aromatic structures.

TABLE 2. Marker Component Average Formula Atomic Numbers, O/C and H/C Ratios, Molecular Weight, Double Bond Equivalents for Plant Biomass Water Extractable Organic Matter, Soil Water Extractable Organic Matter, Mobile Humic Acid, and Calcium Humic Acid Fractions

OM type	number of components	C	H	O	N	S	O/C	H/C	*m/z*	DBE
plant WEOM	772	21.4	32.9	8.01	0.73	0.11	0.40	1.55	431	6.30
soil WEOM	237	20.6	27.3	9.10	0.22	0.20	0.47	1.36	429	8.09
soil MHA	92	26.3	25.6	8.24	0.14	0.12	0.31	0.95	478	14.6
soil CaHA	418	22.4	12.54	7.29	0.32	0.30	0.37	0.64	411	17.3

ponents and the other mostly in the condensed aromatic space (Figure 3d). The CaHA also shows a minor cluster of CHONS components in the carbohydrate region as well. Interestingly, the condensed aromatic components of the CaHA fraction are dominated by N-containing components. This suggests that the humification process that sequesters C by formation of aromatic molecules also sequesters N by incorporation of N to these condensed aromatic OM components. These data support the hypothesis that N-containing components of SOM play a key role in its stabilization through N interactions with mineral surfaces. This is believed to be as result of an amino-N rich inner layer bound directly to mineral surfaces that are part of strongly stabilized aromatic OM structures (*33*). The results shown in Figure 3 clearly show directly that the N-containing OM components are dominant in the aromatic OM fraction. Humification processes can be viewed as a mechanism to sequester and buffer soil N in the more refractory forms of soil OM to protect it from loss through leaching or other mechanisms.

The four samples shown in Figure 3a−d were representative for all samples investigated, as shown by the average % distribution of the components for the four OM types studied (Table S2). The initial phase of the humification process, as represented by the plant WEOM to soil WEOM transition, is marked by increases in the distribution of lignin and carbohydrate classes and decrease in the protein components most likely due to microbial uptake of N-containing OM compounds and release of water-soluble lignin and carbohydrates as the plant biochemical compounds are decomposed. The formation of a moderately refractory MHA fraction is marked by the decrease in the lignin components with increases in the lipid and condensed aromatic components. The formation of the refractory CaHA fraction involves the continued decrease in the lignin and lipid components to increase the condensed aromatic structural compounds. Olk et al. (*12*) used pyrolysis-gas chromatography-mass spectrometry (Py-GC-MS) to determine relative abundances of methoxybenzenes, nonmethoxybenzenes, and nitrogenous and fatty acids/methyl ester compounds in Philippine rice soils. In the dryland soil they investigated, which would be the most similar to the soils used in this study, they also reported higher methoxybenzene (similar to our lignin classification) content in the MHA fraction than in the CaHA. In contrast, they reported a lower content of fatty acid/methyl esters (similar to our lipid classification) in the MHA than in the CaHA fraction. The different lipid (fatty acid/methyl ester) results may be due to the different instrumentation used in the two studies and/or the use of a broader soil set representative of a range of ecoregions in this study as opposed to a single site. The MHA fraction has been reported to be more responsive to cropping management than the more recalcitrant CaHA fraction (*34*). This may be due to the lignin fraction of MHA that is clearly a transitional form of SOM as it continues to undergo humification to the refractory CaHA fraction.

Indicator Species Analysis. The high data density of ultrahigh resolution mass spectrometry provides considerable chemical information, but the analysis and synthesis of that information is challenging with traditional approaches. This suggests the application of chemometric methods to analyze the data to extract ecologically and chemically meaningful information. The set of 37 mass spectra of this study contained 12473 unique components, and the identification of which of those components are likely to be important in tracking how OM is altered through the humification process is difficult. The complexity of this task is highlighted with the plot of the number of components that are found in n samples, where the end member of $n = 1$ is a component that is found in only 1 of the 37 samples and $n = 37$ is a component found in all 37 samples (Figure S1). Forty one percent (5146 of 12473) of the components are found in only one sample and 0.4% (49 of 12473) of the components are found in all 37 samples. The task to link ESI-FT-ICR-MS data to ecosystem and environmental processes will require the identification of specifically which OM components of the thousands of resolved components are involved in the process of interest. Kujawinski et al. (*18*) demonstrated that the ISA statistical method can be used with ESI-FT-ICR-MS data to identify OM marker components that are unique to predefined groups of riverine/estuarine, surface ocean, and deep ocean samples.

Marker components were required to have analysis indicator values greater than or equal to 50, as well as p-values of less than 0.05 for each of the indicator components (*18*). This criterion produced a list of 1519 marker components, representing 12.1% of the total components. Setting the IV threshold to a higher value would have reduced the number of designated marker species. For example, an IV threshold value of greater than or equal to 75 would have reduced the marker species to 208, or 1.7% of the total components. The individual marker components for the four groups using IV of 50 are shown in the Supporting Information, Tables S2−S5. The plant WEOM group had 772 marker components compared to 237 for soil WEOM, 92 for MHA, and 418 for CaHA (Table 2). The large number of markers for the plant WEOM group is not surprising given the large number of biomolecules produced by plants. The two humification endmembers have a higher number of indicator species than the intermediate fractions, which may reflect the transition state of the soil WEOM and MHA where they are actively being transformed to the final stabilized form of SOM. The average elemental, m/z, and DBE data of the marker species show the same trends as the bulk OM patterns but are more strengthened due to the removal of many of the components present across multiple fractions (Tables 1 and 2). The strongest trend in the average elemental data is the decline in the H/C ratio and the corresponding increase in DBE along the plant WEOM to CaHA humification gradient. The van Krevelen plot of the indicator components is shown in Figure 4. The marker components for the plant WEOM are found in the spaces for lipids, proteins, carbohydrates, lignin, and unsaturated hydrocarbons, reflecting the diverse set of water-soluble biomolecules. The soil WEOM OM marker species are predominately concentrated in the higher H/C areas of the lignin space and in the carbohydrate space. The MHA indicator components are clustered in the lower H/C region of the lignin space, and the CaHA are clustered in the

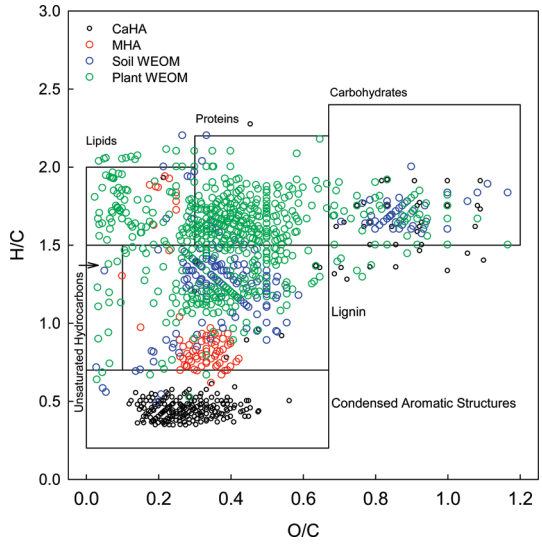

FIGURE 4. van Krevelen diagram of the marker species identified by Indicator Species Analysis. Boxes overlain on the plot indicate major biomolecular compound classes: lipids, proteins, carbohydrates, unsaturated hydrocarbons, lignins, and condensed aromatic structures.

condensed aromatic area. The van Krevelen plot shows that the humification process in soils has a diverse set of "reactants", comprising lipids, proteins, carbohydrates, lignins, and unsaturated hydrocarbons. This ultimately leads to "products" of highly condensed aromatic compounds making up the refractory CaHA fraction.

The ultrahigh resolution ESI-FT-ICR-MS data clearly demonstrate that OM extracted from both plant and soil sources are highly heterogeneous mixtures composed of thousands of components, presumably by hydrophobic interactions and hydrogen bonding. This study demonstrates that the ISA method is an effective statistical tool to determine which components are likely to be involved directly in the processes being investigated. Identification of the marker species combined with graphical analysis tools such as the van Krevelen diagram is effective in reducing the data-rich mass spectra to a more simplified form, while maintaining relevant chemical information. Using this approach, we have shown that humification processes involved in SOM formation are a progression where H:C ratios are steadily reduced from the precursor plant-derived biomolecules to the refractory and stabilized humic acids. The ISA method can also be used to identify specific SOM components to trace throughout an ecosystem pathway, to gain a molecular-scale description and understanding of C dynamics in large-scale ecosystem studies.

Acknowledgments

This project was supported by USDA-National Research Initiative Competitive Grant 2008-35107-04480 from the USDA Cooperative State Research, Education, and Extension Service, and has also been supported by Hatch funds provided by the Maine Agricultural and Forest Experiment Station. We thank Drs. Elizabeth Kujawinski and Krista Longnecker of the Woods Hole Oceanographic Institution for providing and assisting with the MATLAB routine to align the spectra. We also thank Prof. Bruce McCune of Oregon State University for the prompt tweaking of PC-ORD code to handle the large data matrices resulting from ultrahigh resolution mass spectrometry data. This is Maine Agricultural and Forest Experiment Station Journal no. 3152.

Supporting Information Available

Details about the soil analysis and discussion on selected climate and chemical properties of the soils sampled from the field study sites, information about the extraction protocol and EEM/PARAFAC analysis, and elemental formulas for the marker species of the four types of organic matter investigated. This information is available free of charge via the Internet at http://pubs.acs.org/.

Literature Cited

(1) Qualls, R. V.; Richardson, C. J. Phosphorus enrichment affects litter decomposition, immobilization, and soil microbial phosphorus in wetland mesocosms. *Soil Sci. Soc. Am. J.* **2000**, *64* (2), 799–808.

(2) Pham, M. K.; Garnier, J. M. Distribution of trace elements associated with dissolved compounds (< 0.45 um−1 nm) in freshwater using coupled (frontal cascade) ultrafiltration and chromatographic separations. *Environ. Sci. Technol.* **1998**, *32* (4), 440–449.

(3) Taillefert, M.; Lienemann, C. P.; Gaillard, J. F.; Perret, D. Speciation, reactivity, and cycling of Fe and Pb in a meromictic lake. *Geochim. Cosmochim. Acta* **2000**, *64* (2), 169–183.

(4) Horwath, W. H. Carbon cycling and formation of soil organic matter. In *Soil Microbiology, Ecology, and Biochemistry*, 3rd ed.; Paul, E. A., Ed.; Academic Press: New York, 2007; pp 303−339.

(5) Kaiser, K.; Guggenberger, G. The role of DOM sorption to mineral surfaces in the preservation of organic matter in soils. *Org. Geochem.* **2000**, *31* (7−8), 711–725.

(6) Kögel-Knabner, I. A review on the macromolecular organic composition in plant and microbial residues as input to soil. *Soil Biol. Biochem.* **2002**, *34* (2), 139–162.

(7) Jenny, H. *Factors of Soil Formation: A System of Quantitative Pedology*; McGraw-Hill: New York, 1941.

(8) Brady, N. C.; Weil, R. R. *The Nature and Properties of Soils*, 12th ed.; Prentice Hall: Upper Saddle River, NJ, 1999.

(9) Martin, J. P.; Haider, K. Influence of mineral colloids on turnover rates of soil organic carbon. In *Interactions of Soil Minerals with Natural Organics and Microbes*; Huang, P. M.; Schnitzer, M., Eds.; Soil Science Society of America: Madison, WI, 1986; pp 283−304.

(10) Zsolnay, A. Dissolved humus in soil waters. In *Humic Substances in Terrestrial Ecosystems*; Piccolo, A., Ed.; Elsevier: Amsterdam, The Netherlands, 1996; pp 171−223.

(11) Stevenson, F. J. *Humus Chemistry: Genesis, Composition, Reactions*; John Wiley and Sons: New York, 1994.

(12) Olk, D. C.; Dancel, M. C.; Moscoso, E.; Jimenez, R. R.; Dayrit, F. M. Accumulation of lignin residues in organic matter fractions of lowland rice soils: A pyrolysis-GC-MS study. *Soil Sci.* **2002**, *167* (9), 590–606.

(13) Sleighter, R. L.; Hatcher, P. G. The application of electrospray ionization coupled to ultrahigh resolution mass spectrometry for the molecular characterization of natural organic matter. *J. Mass Spectrom.* **2007**, *42* (5), 559–574.

(14) Kujawinski, E. B.; Del Vecchio, R.; Blough, N. V.; Klein, G. C.; Marshall, A. G. Probing molecular-level transformations of dissolved organic matter: Insights from electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Mar. Chem.* **2004**, *92* (1−4), 23–37.

(15) Kim, S.; Kaplan, L. A.; Hatcher, P. G. Biodegradable dissolved organic matter in a temperate and a tropical stream determine from ultra-high resolution mass spectrometry. *Limnol. Oceanogr.* **2006**, *51* (2), 1054–1063.

(16) Gonsior, M.; Peake, B. M.; Cooper, W. T.; Podgorski, D.; D'Andrilli, J.; Copper, W. J. Photochemically induced changes in dissolved organic matter identified by ultrahigh resolution Fourier transform ion cyclotron resonance mass spectrometry. *Environ. Sci. Technol.* **2009**, *43* (3), 698–703.

(17) Kim, S.; Kramer, R. W.; Hatcher, P. G. Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van Krevelen diagram. *Anal. Chem.* **2003**, *75* (20), 5336–5344.

(18) Kujawinski, E. B.; Longnecker, K.; Blough, N. V.; Del Vecchio, F. L.; Kitner, J. B.; Giovannoni, S. J. Identification of possible source markers in marine dissolved organic matter using ultrahigh resolution mass spectrometry. *Geochim. Cosmochim. Acta* **2009**, *73* (15), 4384–4399.

(19) Durfêne, M.; Legendre, P. Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecol. Monogr.* **1997**, *67* (3), 345–366.

(20) Brown, T. L.; Rice, J. A. Effect of experimental parameters on the ESI FT-ICR mass spectrum of fulvic acid. *Anal. Chem.* **2000**, *72* (2), 384–390.

(21) Rostad, C. E.; Leenheer, J. A. Factors that affect molecular weight distribution of Suwannee river fulvic acid as determined by electrospray ionization/mass spectrometry. *Anal. Chim. Acta* **2004**, *523* (2), 269–278.
(22) Ohno, T.; Bro, R. Dissolved organic matter characterization using multi-way spectral decomposition of fluorescence landscapes. *Soil Sci. Soc. Am. J.* **2006**, *70*, 2028–2037.
(23) Djikanovic, D.; Kalauzi, A.; Radotic, K.; Lapierre, C.; Jeremic, M. Deconvolution of lignin fluorescence spectra: A contribution to the comparative structural studies of lignins. *Russian J. Phys. Chem. A* **2007**, *81* (9), 1425–1428.
(24) Lundquist, H.; Josefsson, B.; Nyquist, G. Analysis of lignin products by fluorescence spectroscopy. *Holzforschung* **1978**, *32* (1), 27–32.
(25) Kowalczuk, P.; Durako, M. J.; Young, H.; Kahn, A. E.; Cooper, W. J.; Gonsior, M. Characterization of dissolved organic matter fluorescence in the South Atlantic Bight with use of PARAFAC model: Interannual variability. *Mar. Chem.* **2009**, *113*, 182–196.
(26) Sleighter, R. L.; Liu, Z.; Xue, J.; Hatcher, P. G. Multivariate statistical approaches for the characterization of dissolved organic matter analyzed by ultrahigh resolution mass spectrometry. *Environ. Sci. Technol.* **2010**, *44* (19), 7576–7582.
(27) Koch, B. P.; Witt, M.; Engbrodt, R.; Dittmar, T.; Kattner, G. Molecular formulae of marine and terrigenous dissolved organic matter detected by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Geochim. Cosmochim. Acta* **2005**, *69* (13), 3299–3308.
(28) Kim, S.; Kaplan, L. A.; Benner, R.; Hatcher, P. G. Hydrogen-deficient molecules in natural riverine water samples- evidence for the existence of black carbon in DOM. *Mar. Chem.* **2004**, *92* (1), 225–234.
(29) Sutton, R.; Sposito, G. Molecular structure in soil humic substances: The new view. *Environ. Sci. Technol.* **2005**, *39* (23), 9009–9015.
(30) Hertkorn, N.; Frommberger, M.; Witt, M.; Koch, B. P.; Schmitt-Kopplin, Ph.; Perdue, E. M. Natural organic matter and the event horizon of mass spectrometry. *Anal. Chem.* **2008**, *80* (23), 8908–8919.
(31) Trembay, L. B.; Dittmar, T.; Marshall, A. G.; Cooper, W. J.; Cooper, W. T. Molecular characterization of dissolved organic matter in a North Brazilian mangrove porewater and mangrove-fringed estuaries by ultrahigh resolution Fourier transform-ion cyclotron resonance mass spectrometry and excitation/emission spectroscopy. *Mar. Chem.* **2007**, *105* (1−2), 15–29.
(32) Hockaday, W. C.; Purcell, J. M.; Marshal, A. G.; Baldock, J. A.; Hatcher, P. G. Electrospray and photoionization mass spectrometry for the characterization of organic matter in natural waters: A qualitative assessment. *Limnol. Oceanogr.-Meth.* **2009**, *7*, 81–95.
(33) Hsu, P.-H.; Hatcher, P. G. New evidence for covalent coupling of peptides to humic acids based on 2D NMR spectroscopy: A means for preservation. *Geochim. Cosmochim. Acta* **2005**, *69*, 4521–4533.
(34) Olk, D. C. A chemical fractionation for structure-function relations of soil organic matter in nutrient cycling. *Soil Sci. Soc. Am. J.* **2006**, *70* (3), 1013–1022.

ES101089T