1 Title of the manuscript:

2 Potential origin and formation for molecular components of humic acids in soils

3

4 Author names and affiliations:

5 Nicole DiDonato^{a1} (ndidonat@odu.edu), Hongmei Chen^a (hchen015@odu.edu), Derek
6 Waggoner^a (DWaggoner@odu.edu), Patrick G. Hatcher^{a*} (phatcher@odu.edu)

7 ^a*Department of Chemistry & Biochemistry, College of Sciences, Old Dominion University,*
8 *Norfolk, VA 23529 USA*

9 ¹contact author prior to publication

10 *corresponding author post-publication

11 Patrick G. Hatcher

12 Present address: *Department of Chemistry & Biochemistry, College of Sciences, Old Dominion*
13 *University, Norfolk, VA 23529 USA*

14 Tel: +1.757.683.6537

15 Fax: +1.757.683.5310

16 E-mail address: phatcher@odu.edu

17

18

Abstract

Soil humic acids are the base soluble/acid insoluble organic components of soil organic matter. Most of what we know about humic acids comes from studies of their bulk molecular properties or analysis of individual fractions after extraction from soils. This work attempts to better define humic acids and explain similarities and differences for several soils varying in degrees of humification using advanced molecular level techniques. Our investigation using electrospray ionization coupled to Fourier transform ion cyclotron resonance mass spectrometry (ESI-FTICR-MS) and nuclear magnetic resonance spectroscopy (NMR) has given new insight into the distinctive molecular characteristics of humic acids which suggest a possible pathway for their formation. Humic acids from various ecosystems, climate regions and soil textural classes are distinguished by the presence of three predominant molecular components: lignin-like molecules, carboxyl-containing aliphatic molecules and condensed aromatic molecules that bear similarity to black carbon. Results show that humification may be linked to the relative abundance of these three types of molecules as well as the relative abundance of carboxyl groups in each molecular type. This work also demonstrates evidence for lignin as the primary source of soil organic matter, particularly condensed aromatic organic molecules often categorized as black carbon and is the first report of the non-pyrogenic source for these compounds in soils. We also suggest that much of the carboxyl-containing aliphatic molecules are sourced from lignin.

Abbreviations

ESI-FTICR-MS, electrospray ionization coupled to Fourier transform ion cyclotron resonance mass spectrometry; NMR, Nuclear Magnetic Resonance Spectroscopy; DBE, double bond equivalents; CRAM, carboxyl-rich alicyclic molecules; CCAM: carboxyl containing aliphatic molecules; BC, Black carbon; CPMAS, cross-polarization magic angle spinning; DPMAS, direct polarization magic angle spinning; KMD, Kendrick mass defect; Ai_{mod} , modified aromaticity index

1. INTRODUCTION

Humification of soil organic matter has traditionally been defined as the process by which plant and microbial debris are transformed into humic substances (Guggenberger, 2005; Sollins et al., 2007; Stevenson, 1994). Humic substances themselves are only loosely defined as the significantly altered macromolecules remaining in soils that are not composed of easily recognizable known biomolecules (carbohydrates, lignin, proteins, amino-sugars, etc.). These have further been defined operationally into three categories: base-soluble/acid-insoluble humic acid, acid/base soluble fulvic acid and insoluble humin.

Proposed pathways for the formation of humic substances include the lignin and lignin decomposition theories, the lignin-polyphenol theory as well as the melanoidin pathway (Stevenson, 1994). It is generally accepted that a combination of several of these pathways with some modifications may be responsible for producing humic substances. Also, it is well known that physical and chemical conditions such as climate (temperature and moisture), redox conditions, pH as well as the amount and chemical compositions of organic matter source inputs determine which and how far these humification processes progress. In this paper, we refer to more 'humified' samples as those subject to these processes to a greater extent due to environmental conditions or for a longer period of time. Most commonly, though not exclusively, the extent of humification can be illustrated through changes in organic matter contained within soil horizons at increasing depth of the soil column; more 'humified', maturely decomposed material tending to be found at depth and less humified, less decomposed material at the surface (Kögel-Knabner et al., 1991). It is generally agreed that more humified organic matter is present in highly aerobic soils than in soils that experience a lower degree of aerobic decomposition (Ikeya et al., 2013).

For well over 100 years attempts have been made to structurally define the nature of humic substances and numerous models have been developed to apply a global definition of the structural entities that might constitute humic substances. Few of these models incorporate the recent views from ultrahigh resolution mass spectrometry that humic substances are generally low molecular weight (~500 Da) but extremely varied in structure (Kramer et al., 2001; Sutton and Sposito, 2005). The general picture that emerges from all the previous work and this recent work is that humic substances comprise molecules containing varying proportions of aromatic structures, aliphatic structures, proteinaceous materials, sugars, and lipids. The aromatic and aliphatic structures appear to harbor functional groups that define their solubility or insolubility in alkali and dilute acids.

The recent vision of humic substances derived from ultrahigh resolution mass spectrometry seems to suggest that thousands of molecular components exist and these molecules have some common features that allow us to class them into several types of molecules defined by the van Krevelen diagram which plots molar H/C ratios vs molar O/C ratios (van Krevelen, 1950). This type of plot was first used by Kim et al. (2003) for ultrahigh resolution mass spectral data

obtained by electrospray ionization coupled to Fourier transform ion cyclotron resonance mass spectrometry (ESI-FTICR-MS). Since this first paper, hundreds of studies have employed this approach to characterize humic substances (Nebbioso and Piccolo, 2013; Sleighter and Hatcher, 2007). What is particularly salient to our understanding of humic substances is that van Krevelen plots have defined clusters of molecules that differentiate various molecular entities that serve as a basis for the current focus. The first of these represents molecules with a molecular similarity to fragments of lignin (200-800 Da), an important biopolymer component of terrestrial plants. Similarly, other regions of the van Krevelen diagram represent molecules showing a similarity to proteinaceous materials, tannins, lipids, and carbohydrates, all of which have a derivation from well-known plant biopolymers. Two other regions of van Krevelen space have been noted. One of these is a region of condensed aromatic molecules commonly associated with pyrolytic destruction of organic compounds (Hockaday et al., 2006; Ikeya et al., 2013; Kramer et al., 2004). These appear to be sufficiently adorned with oxygenated functional groups (OH, COOH, CHO) as well as N or S-containing groups to become soluble and ionizable for ESI-FTICR-MS analysis. The other region spans the van Krevelen diagram between H/C ratios of 2.0 and 0.85 and appear to be mostly aliphatic in nature (Ohno et al., 2010). However, the number of double-bond equivalents (DBE) and oxygenated functional groups associated with these molecules suggests that they include a class of molecules named by Hertkorn et al. (2006) as CRAM (carboxyl-rich alicyclic molecules). The CRAM notation was originally used for marine-derived dissolved organic matter but it is clear that, in terrestrial dissolved and sedimentary organic matter, one observes aliphatic molecules that fit the CRAM definition in that they have multiple double-bond equivalents and contain carboxyl groups (Lam et al., 2007). In humic acids from terrestrial sources, since these molecules plot in a region beyond the CRAM region as defined by Hertkorn, we refer to them here simply as carboxyl containing aliphatic molecules (CCAM). These molecules appear to be ubiquitous components of humic substances suggesting that they might be microbial biomolecules altered by environmental processes (Nebbioso et al., 2014).

It has become clear from recent work by our group (Ikeya et al., 2013; Kramer et al., 2004; Ohno et al., 2010) and others (Grinhut et al., 2011) that humic substances in soils, humic acids in particular, show ultrahigh resolution MS data that display van Krevelen clusters of points in three principal regions: the lignin-like, the carboxyl containing aliphatic molecules (CCAM), and the condensed aromatic, or black carbon-like (BC-like) region. To demonstrate, Figure 1 shows a van Krevelen diagram for a soil humic acid whose partial mass spectral data were published by Ohno et al. (2010), one that appears to be representative for humic acids from numerous soils distributed across the USA. The respective regions for the three classes of molecules are shown.

The current study examines humic acids from additional soil samples to further investigate the ubiquitous presence of molecular formulas plotting in the three defined regions. In addition we provide an explanation for the formation of these molecules that introduces a new perspective of the humification process. Two of the soils were collected from a single site, however, they vary

in degree of humification because one is developed on a poorly drained substrate and the other is well drained such that the organic matter accumulates under continuously oxidizing conditions. The former poorly drained soil is inundated during the wet season and is also adjacent to a small stream. Accordingly, this soil is not as humified. Other factors (pH, organic matter input, occlusion and stabilization by soil minerals/clays, etc.) are also thought to influence the extent of humification, however, in the case of these two samples it is expected that variations in water content, aeration and sunlight play the most dominant role in this process. Our work utilizes advanced analytical techniques such as ESI-FTICR-MS and solid state NMR to more completely characterize these humic acids at the molecular level. The ESI-FTICR-MS mass spectra of a variety of other soil humic acids are also evaluated for comparison. While it is rather straightforward to explain the source of the lignin-like molecules, we propose a new process for the formation of the other two classes of molecules that compose humic acids. Because these appear to be ubiquitous across soils representing a diversity of soil environments, the process we suggest may be a global one.

2. MATERIALS AND METHODS

2.1 Humic acid samples

Georgia humic acids were extracted from soils using 0.5 M NaOH followed by treatment with a Dowex™ ion-exchange resin to remove sodium ions. The resin was repeatedly cleaned with MilliQ water to neutrality. Humic acids were precipitated by 6 M HCl at pH 2 and freeze dried. The humic acids from Ohno et al. (2010) were treated by a process described by Olk et al. (2002) which separates two types of humic acids, those extracted without soil treatment and those extracted after acidification of the soil to release Ca-bound humic acids. We examined only the latter as they appear to be the most highly humified. The Tsubame humic acid from a buried humic layer of a highly humified forest soil in Japan was prepared as described by Ikeya et al. (2013). The Georgia soil humic acids (GA-1 and GA-2) were taken from the surface layer (top 10 cm) of a swampy hardwood forest soil and a nearby grassy, open field, respectively, in Kingsland, GA. Armadale humic acids were obtained from the surface A_o horizon of a spodosol soil from a deciduous forest in the Armadale region of Prince Edward Island, Canada. These samples were provided to us by Dr. Morris Schnitzer and prior NMR studies have been published (Chen and Schnitzer, 1978; Hatcher and Clifford, 1994; Kramer et al., 2001; Schulten and Schnitzer, 1992). Available soil properties and characteristics are tabulated in Table S1.

2.2 ESI-FTICR-MS

Georgia and Armadale humic acids were dried to remove water, weighed and re-dissolved in 0.01 M NaOH overnight (~12 hours) under argon headspace. They were then batch treated with a Dowex™ 50WX8-100 ion-exchange resin at a 3:1 v/v ratio and shaken for 1 hour to remove Na and other cations to reduce the formation of salts upon ionization. Analysis was conducted in the negative ion mode on a Bruker Daltonics 12 Tesla Apex Qe FTICR-MS instrument equipped

with an Apollo II ESI source and housed in the College of Sciences Major Instrumentation Cluster at Old Dominion University. To improve ionization efficiency, samples were mixed 1:1 with methanol just prior to injection for a final concentration of approximately 50 ppm. A dichloromethane rinse was performed to remove organics sorbed to the Dowex resin and these samples were run separately from the aqueous samples after mixing 1:1 with methanol prior to injection. Ions were accumulated in the hexapole for 2 seconds before transfer to the ICR cell where a 4 megaword time domain was used to obtain 300 scans co-added in broadband mode using an m/z range of 200-1200. Spectra were externally calibrated to polyethylene glycol and internally calibrated to a fatty acid series common to natural organic matter (Sleighter et al., 2008). Empirical molecular formulas were assigned in the mass range from 200-800 m/z using an in-house Matlab code (The MathWorks, Inc., Natick, MA) according to the following criteria: $^{12}\text{C}_{2-50}$, $^1\text{H}_{5-100}$, $^{14}\text{N}_{0-6}$, $^{16}\text{O}_{1-30}$, $^{32}\text{S}_{0-2}$, and $^{31}\text{P}_{0-2}$ within an error of 1 ppm, and using the rules outlined by Stubbins et al. (2010). For this study, only CHO formulas were investigated, and they represent a majority (45-73%) of the total number of assigned formulas (760-1206). CHO formulas from the aqueous samples were combined with those assigned for their respective DCM extract to make a combined data set for each humic acid. Valentine and Catlin humic acids were analyzed according to Ohno et al. (2010) and Tsubame humic acids were analyzed according to Ikeya et al. (2013).

Molecular formulas were investigated using the van Krevelen diagram and categorized into three regions based on where the formulas plotted: lignin-like ($0.2 \leq \text{O/C} \leq 0.6$, $0.6 \leq \text{H/C} \leq 1.2$), carboxyl containing aliphatic molecules (CCAM) ($0.85 \leq \text{H/C} \leq 2$, $\text{O/C} \leq 0.4$) and condensed aromatic (modified aromaticity index, $\text{Ai}_{\text{mod}} > 0.67$) (Koch and Dittmar, 2006), where:

$$\text{Ai}_{\text{mod}} = \frac{1 + \text{C} - 0.5 * \text{O} - \text{S} - 0.5 * \text{H}}{\text{C} - 0.5 * \text{O} - \text{S} - \text{N} - \text{P}} \quad (1)$$

It is important to note that there is some overlap between the CCAM-like and lignin-like regions, such that a small portion of molecular formulas plotting in both of these regions with $0.2 < \text{O/C} < 0.4$, $0.8 > \text{H/C} > 1.2$ could represent either or both of the CCAM or lignin-like categories of molecular types based purely on their elemental ratios. For this reason, the fraction of formulas plotting in each region does not sum to unity.

The formulas were also examined according to their Kendrick mass defect (KMD) for the carboxyl group (COO), which is calculated as follows (Ikeya et al., 2013; Kramer et al., 2004):

$$\text{KMD} = \text{Kendrick mass (COO)} - \text{Nominal Mass} \quad (2)$$

$$\text{Kendrick mass (COO)} = \text{Exact Mass} * (\text{Nominal mass COO} / \text{Mass COO}) \quad (3)$$

2.3 NMR

Solid State ^{13}C NMR spectra were obtained on a Bruker 400 MHz Avance II spectrometer equipped with a 4 mm solid state MAS probe. Samples were packed into an 80 μL zirconia rotor

and sealed with a Kel-F cap. Samples were spun at 11K Hz to avoid operating at the threshold of stability for the rotor throughout the long acquisition time of the DPMAS experiment. A small error may be associated with spinning sidebands at this frequency, however we do not expect this to be much above the noise or exceed 1% of the signal intensity. 6400 scans were acquired during cross-polarization magic angle spinning (CPMAS) and 2000 scans were acquired during direct polarization magic angle spinning (DPMAS). A 1 second recycle delay and contact time of 1 ms was used for CPMAS. Contact times were varied from 1 ms to 10 ms to find the most appropriate contact time for magnification transfer. For DPMAS, the recycle delay was set to 50 sec to obtain relaxation of all carbons. Glycine was used as an external standard calibrated to C-2 at 42.56 ppm. Integrations were performed using the Bruker Topspin software.

2.7 Elemental analysis

The elemental composition of the Georgia humic acids was evaluated using a FlashEA 1112 elemental analyzer containing a CHN column. Samples were prepared in triplicate and calibrated to an acetanilide standard curve.

3. RESULTS

3.1 Composition of Georgia Humic Acids

3.1.1 ESI-FTICR-MS

Van Krevelen diagrams for mass spectral data from the two soil humic acids from the same region in Kingsland, GA (GA-1 and GA-2), but representing two different stages of humification are shown in Figure 2. Figure 2a shows the ESI-FTICR-MS-generated elemental data for humic acid GA-1, obtained from a poorly drained soil sampled from a deciduous forest swamp. The standing water and low oxygen content limits humification of organic matter inputs such as plant debris. In contrast, humic acids were also sampled from a nearby open field (GA-2) with grass cover. Humic acids here are exposed to more oxygen and sunlight. Regular drying of the soil allows organic matter inputs, mostly grasses, to decompose rapidly in the aerobic environment. Humic acid contents in soils can vary, and although we did not measure recoveries in this study, they can account for up to 40-50% of the organic carbon in soils.

What is immediately apparent from the data in the van Krevelen diagrams is that points appear to be concentrated in the three regions defined previously to be ubiquitously present in soils from various other locations (Ohno et al., 2010). Points in the lignin-like region represent molecules derived from the macromolecular lignin polymer that has undergone microbial degradation to small molecules and has most likely been oxidized such that it becomes soluble in base to be included as humic acids. It should be noted that other aromatic biopolymers such as condensed tannins may also plot in this region however we do not suggest that these molecular types make a significant contribution to these samples. They constitute a much lower percentage of total fixed carbon in land ecosystems than lignin, which is considered second only to cellulose as the most

abundant constituent of plant biomass and makes up about 30% of the organic carbon in the biosphere (Boerjan et al., 2003; Ruiz-Dueñas and Martínez, 2009). Hydrolysable tannins, which plot in a region with O/C as high as 1, are either absent or poorly represented in the humic acid ESI-FTICR-MS data, however the data does not preclude their presence within these samples. Points in the condensed aromatic region have been suggested to originate from pyrolytic processes and to be mainly substituted by carboxyl groups (Ikeya et al., 2013; Kramer et al., 2004; Ohno et al., 2010). Points in the aliphatic CRAM-like region have been suggested to derive from microbial lipids or plant terpenoids (Simpson et al., 2011; Singer et al., 2012).

Fewer molecular formulas in the lignin region of GA-1 (36% vs 42% of total peaks in GA-2, Table 2) may be due to decreased input of these molecules to soil humic acid as a result of decreased humification of the intact plant biopolymers (Kurbatov, 1968). Microbiological activity needed for chemical breakdown of these organic materials may be inhibited by the hydro-topography of this area. However, once exposed by enzymatic attack and increased humification, more molecular formulas, and a greater variety that are similar to lignin-like degradation products can be detected in humic acids as we observe in Figure 2b. With increased humification and the ensuing increased enzymatic attack by microorganisms, molecular formulas become more pronounced in the condensed aromatic region as well. This is observable in Figure 2 and Table 2 by comparing the relative number of peaks plotting in the condensed aromatic region (BC-like region; 25% vs 9% of total peaks for GA-2 and GA-1, respectively). A similar trend has been noted for humic acids from Japanese soils (Ikeya et al., 2013).

ESI-FTICR-MS mass spectra have been shown to be reproducible (Sleighter et al., 2012) and peak magnitudes may be considered semi-quantitative where response factors can be established for a majority of the molecules contained within a sample (Kamga et al., 2014). The technique has been used extensively for characterizing samples and comparing compositions based on formula presence/absence, relative peak magnitude and the use of van Krevelen diagrams (Antony et al., 2014; Hur et al., 2010; Kim et al., 2003; Koch et al., 2008; Mesfioui et al., 2012; Reemtsma, 2009; Sleighter and Hatcher, 2008; Sleighter et al., 2009; Stubbins et al., 2010). Future work may seek to quantify absolute concentrations for individual components, but ESI-FTICR-MS is useful for our purposes to evaluate the relative abundance of compounds within a given sample and among several samples based on relative percentages of peak numbers and summed peak intensities for the van Krevelen regions of interest as shown in Table 2. NMR results also provide for a more quantitative evaluation of these samples.

3.1.2 NMR

Lignin-like compounds are also apparent from the solid state NMR spectra of GA-1 and GA-2 humic acids shown in Figure 3. Peaks characteristic of both syringyl (153 ppm) and guaiacyl (147 ppm) lignin are present, which is expected for soils from hardwood forests containing both G and S lignin such as GA-1, as well as grasslands (GA-2) which contain all three H, G and S lignin units (Koenig et al., 2010). While the peak at 147 ppm is conspicuous in both spectra, its

intensity is elevated with respect to the syringyl peak in the grassland soil humic acid (GA-2). This could represent the relative contribution of lignin monomers to the soil based on the composition in the individual species present at the site, however this peak is also assigned to non-etherified guaiacyl and syringyl units of lignin and may also indicate relatively more depolymerized lignin in the GA-2 sample (Chefetz et al., 2000; Koenig et al., 2010).

Lignin ring methoxy (56 ppm) and aryl-O carbons (150 ppm) are also prevalent in these spectra (Hatcher, 1988). The ratio of the area of these peaks (0.58) in the DPMAS spectra for the forested area matches well with the expected lignin monomer compositions for this ecosystem, assuming angiosperms are composed of an equal mixture of guaiacyl and syringyl monomers to obtain an average aryl-methoxy to aryl-O ratio of 0.58.

The DPMAS spectra also reveal the high degree of aromaticity of these samples. Comparing the aryl-O carbon peak area to the aromatic carbon peak area for these samples we notice a decrease in this ratio for GA-2 in comparison to GA-1 (Table 3). This is in line with relatively less O-substituted rings and potentially more humified material in GA-2.

The relative amounts of non-protonated aromatic carbons in each of the samples is also striking when comparing the DPMAS spectra with the corresponding CPMAS spectra for each sample. Because direct polarization of the carbon nuclei magnetizes all carbons equally, the spectra obtained are quantitative. In contrast, CPMAS preferentially transfers magnetization to protonated carbons, and so we can take the difference between these two spectra to estimate the amount of non-protonated aromatic, or condensed aromatic carbon in the samples. Sp³ hybridized carbons are least affected by the difference between these two pulse programs, and so we can compare the two spectra by normalizing the integrals of each spectral region to the aliphatic region (0-45 ppm). This reveals a 16% enhancement in the area of the aromatic region (120-140 ppm) for GA-2 with direct polarization through the carbon nuclei (Table 4). In contrast, the aromatic region in the less humified sample (GA-1) that has been preserved in a poorly drained forest area contained about 12% non-protonated carbon when normalized to the aliphatic region of the CP spectra (Table 4). This is also in agreement with the ESI-FTICR-MS data, which showed a lower fraction of condensed aromatic formula for this sample than GA-2 as discussed previously in section 3.1.1. Although grasslands are prone to fire events which can contribute condensed aromatics to the soils, GA-2 samples were taken in a swampy region of southern Georgia where fires are less common. A search of the USGS Federal Fire Occurrence database, which includes data from the 1980s, showed no records of fire events in this region.

Signals from carbohydrates or alkyl-O carbons (72 ppm) are relatively weak in comparison and may also indicate the degree of humification of these samples. There is relatively less signal from carbohydrate carbons in GA-2. Likewise, the increased intensity of peaks at 168-172 (carboxyl) relative to the aliphatic region and reduced intensity of the peak at 196 (carbonyl) is in line with the sample from GA-2 containing more oxidized lignin and a higher humification potential. This is also supported by the higher percentage of molecular formulas observed by

ESI-FTICR-MS in GA-2 (69% vs. 55%) to be contained within a COO Kendrick mass defect series as explained in the next section.

3.2 Comparison of Humic Acids

3.2.1 Formula Types

Soil humic acids from several other soils were also investigated in this study. The van Krevelen diagram for mass spectral data of humic acids from the Catlin soil of Ohno et al. (2010) and the Japanese soil of Ikeya et al. (2013) are depicted in Figure 4a and c. The formulas fall clearly into two predominant regions consisting of condensed aromatic molecules as well as high H/C, low O/C aliphatic CCAM molecules. Few points are observed in the lignin-like region. In contrast, the spectral data for humic acids from a spodosol soil (Armadale soil) in Figure. 4b show a less dramatic separation of these regions, with relatively more molecular formula plotting in the lignin-like region and relatively fewer condensed aromatic molecules.

From the mass spectral observations made for the humic acids, we can readily discern a relationship based on degree of humification. The less humified samples contain relatively more components having a relationship to lignin (Ikeya et al., 2015; Kögel-Knabner et al., 1991). The more humified samples appear to mainly comprise molecules that are either condensed aromatic or are CCAM.

The range of this phenomenon is also apparent when we graph the distribution of molecular formulas in eight additional humic acid samples from the Ohno et al. (2010) study (Figure 5). These samples represent humic acids from an assortment of soil types and climate regions and should represent the diversity of soil environments across the United States. They are also considered to contain more humified organic material as a result of the isolation procedure in which more mobile humic acids were removed. These humic acids were distinguished from their corresponding soil and plant biomass aqueous extracts by a decrease in lignin-like components and increase in both “lipid-like” and condensed aromatic components (Ohno et al., 2010). The trend is towards a collection of formulas and ion intensities primarily in the CCAM and condensed aromatic regions, with the remaining formulas plotting in the lignin-like region of the van Krevelen diagram, as can be seen in Figure 5. The average percentage of formulas and ion intensities, respectively, in the three regions were found to be: $43\% \pm 12\%$ (76% intensity $\pm 15\%$) for CCAM, $48\% \pm 13\%$ (22% intensity $\pm 14\%$) for condensed aromatic and $9\% \pm 6\%$ ($2\% \pm 2\%$ intensity) for lignin-like regions.

3.2.2 Carboxyl groups

While it is generally well recognized that carboxyl groups are the fundamental constituent functionalities that define humic acids, and the NMR data bear this out, we can query the elemental formula makeup of the mass spectral dataset for an indication that carboxyl groups may be components of the molecular formulas. This is accomplished by a Kendrick mass defect

plot that manipulates the formula composition to align elemental formulas that differ by an exact mass of the COO group (Ikeya et al., 2013; Kramer et al., 2004). If we consider the five humic acids representing varying levels of humification (GA-1, Armadale, GA-2, Valentine and Catlin), approximately 53-81% of the molecular formulas assigned for these samples were found to be present as part of a COO KMD series of 2 or more (Figure 6), where these series extend up to six molecular formulas.

For example, Figure 7a shows the formulas in the van Krevelen diagram for humic acids from soils at the USDA site in Catlin, IL (blue, also Fig. 4a) which are contained within a COO KMD series (green, Fig. 7a). These series include up to five molecular formula as shown in Figure 7b and include 81% of the CHO formulas assigned for this sample.

When comparing the percentage of all formulas contained within a COO series among samples, those samples considered to have a higher degree of humification (Valentine, Catlin, GA-2) in general contained more formulas within a COO KMD series (75%, 81% and 70% respectively) in comparison to less humified samples (53 and 55% for the Armadale and GA-1, respectively) (Figure 6). This extends the relationship established above. The more humified samples are composed of relatively more molecular formulas in the CCAM and condensed aromatic regions and also contain relatively more carboxylated molecular formulas than the less humified samples.

In addition, by comparing COO containing formulas for individual regions (lignin-like, condensed aromatic and CCAM) of these samples we find that molecular formulas in the regions representing more humified material (condensed aromatic and CCAM molecular formulas) in general contain relatively more COO containing formulas than the lignin-like region for each sample (Figure 8). The Valentine, NE and GA-2 humic acids are the exception, with 7 and 9% less COO KMD series formulas in the respective CCAM regions relative to the lignin regions. However, a greater percentage of COO KMD series formulas are present in the condensed aromatic regions for both samples as compared to the lignin regions (95% and 85%, respectively). Of the humic acids studied here, these two also have the lowest percentages of all formulas that fall in the CCAM region (41% and 50% respectively).

Overall, molecular formulas from the CCAM regions for these samples contained 54-82% of formulas that are part of a COO Kendrick mass defect series (Figure 8). Formulas included in a COO KMD series also accounted for 40-95% of the formulas in the condensed aromatic regions of these samples. In comparison, the lignin-like region contained relatively less, 14-65%, of formulas that were included in COO KMD series. Considering that carboxylic acids are known to be produced during lignin oxidation (Chen et al., 1983; Leonowicz et al., 2001), such a high proportion of COO KMD series formulas within the CCAM and condensed aromatic regions in comparison to the lignin region is further evidence that these regions may represent secondary, or more 'humified' molecules.

388 Soil humic acids from varying locations, soil types and degrees of humification are characterized
389 by a common set of molecular formulas and structural entities whose presence has been debated
390 to either derive from lignin and modified lignin, sugars and proteins, fungal or microbial
391 phenols, or a combination of all these (Stevenson, 1994). Some recent studies in our group (Chen
392 et al., 2014) indicate that radical polymerization reactions could play an important role in altering
393 natural organic matter that is dominated by terrestrial lignin-derived organic matter. Chen et al.
394 have shown that photo-irradiation of dissolved organic matter (DOM) from the Great Dismal
395 Swamp, Virginia, produces new particulate, base-soluble molecules of two distinct structural
396 classes: high H/C, low O/C aliphatic (CCAM formulas) and low H/C aromatic and condensed
397 aromatic molecular formulas with hydroxy and carboxyl substituents. The aromatic formulas,
398 similar to black carbon from combustion sources, represent a previously unrecognized source of
399 condensed aromatics and match formulas found in peat humic acids from the Dismal Swamp.
400 The DOM from the Dismal Swamp is dominated by molecules derived predominantly from
401 lignin (Hartman et al., 2015). Although primarily light-induced in the Chen et al. study,
402 hydroxyl radicals responsible for catalyzing this oxidation are also present in soil systems. Here,
403 fungal enzymes are one source of hydroxyl radicals (Carlile et al., 2001), others include humic
404 substances themselves (Page et al., 2013; Page et al., 2012), and we hypothesize that reactions
405 similar to those discovered by Chen et al. likely take place.

406 The details of lignin biodegradation pathways in soils can vary by organism and wood type and
407 are not well understood, except for a few microbes and enzymes. White rot fungi are known to
408 utilize lignin and manganese peroxidases to degrade lignin to CO₂ via several reactions, all
409 relying on hydrogen peroxide to incorporate oxygen. Further, ring hydroxylation and opening by
410 enzymatic hydroxyl radicals is known to produce unsaturated aliphatic carboxylic acids and
411 hydroxylated muconic acids (Higuchi, 2004). Another recent study by our group has
412 demonstrated similar results using Fenton reaction chemistry to produce hydroxyl radicals for the
413 degradation of a lignin-rich wood extract (Waggoner et al., 2015). As with the study by Chen et
414 al. the ESI-FTICR-MS results showed not only the generation of new aliphatic molecules
415 containing carboxylic acids, which fall in the CCAM region of the van Krevelen, but new
416 condensed aromatic molecules as well.

417 Accordingly, we propose that photo-, abiotically- or microbially-generated hydroxyl radicals in
418 soils are responsible for transforming the materials supplied to soil as fresh organic matter,
419 mainly lignin, to the molecules observed in ESI-FTICR-MS data. When plotted on the van
420 Krevelen diagram, the H/C and O/C ratios of molecular formulas from humic acids predictably
421 plot in the same regions as the newly produced formulas discovered by Chen, et al. In fact, a
422 portion of these molecular formulas were found to have the same exact composition as those
423 identified by Chen et al. Figure 9 illustrates the photo-produced molecular formulas (red) in
424 common with the humic acid formulas from the Valentine, NE site (blue) from Figure 1. These

formulas represent approximately 16% of the humic acid CHO molecular formulas assigned for this sample.

Based on these observations, we have defined the three regions of interest for the humification process in soils as follows. (1) The region for lignin-like compounds ($0.2 \leq O/C \leq 0.6$, $0.6 \leq H/C \leq 1.2$) we propose to be the primary substrate for subsequent reactions with hydroxyl radicals in which ring opening, condensation and transformation into two distinct types of molecules occurs. These are (2) condensed aromatics ($Ai_{mod} > 0.67$) and (3) carboxyl containing aliphatic molecules (CCAM) ($0.85 \leq H/C \leq 2$, $O/C \leq 0.4$) reminiscent of carboxyl-rich alicyclic molecules (CRAM) described by Hertkorn et al. (2006).

Based on KMD COO series, carboxylic acids may functionalize the molecules that plot in these regions and we suggest this could be the result of two processes. First, the initial oxidation of lignin is known to produce carboxylic acids (Higuchi, 2004) (see also Figure 2 of Waggoner et al. 2015). These molecules would plot with higher O/C in the lignin-like or CCAM-like region of the VK diagram, as was shown in the Waggoner et al. study. Secondly, decarboxylation may occur through the loss of CO₂ during radical propagation of muconic acid-like molecules (Figure 3 Scheme, Waggoner et al. 2015) and these molecules would plot in the lower O/C and higher H/C CCAM-like region. Decarboxylation may also occur during intermolecular cyclization and radical condensation reactions (Figures 4 and 7, Waggoner et al. 2015), with these molecules plotting in the condensed aromatic or possibly the CCAM-like regions. Depending on the extent of the reactions and intermediates formed, the presence of COO KMD series within CCAM like molecules could be evidence for the first process of carboxylic acid formation as well as the second process of decarboxylation, or both. The presence of molecules differing only by the number of carboxyl groups implies these molecules have the same core molecular formula and could be reactants or products of these processes. However, the direction of the process (carboxylation or decarboxylation) cannot be discerned from this information alone.

The modified aromaticity index (Ai_{mod}) is used here as a conservative approach to delineate the condensed aromatic region since it assumes half of the double bonds formed between C and O do not contribute to aromaticity, ring formation or condensation (Koch and Dittmar, 2006). While the actual aromaticity in a molecule can be higher, an $Ai_{mod} \geq 0.67$ provides a minimum threshold for the presence of condensed aromatic structures.

The average DBE for molecular formulas of the humic acids in this study categorized as CCAM ranges from 3.6 to 9.4, and the average Ai_{mod} ranges from 0.10-0.27, within the range for unsaturated and/or alicyclic molecules. There is evidence for the biological origin of some of these molecules based on the even over odd carbon number predominance of molecular formulas with low DBE, which suggests contributions from plant waxes (Gupta, 2014). The intensity-weighted bubble plot of DBE vs. carbon number shown in Figure 10 illustrates this most clearly for molecular formulas from C₁₄ to C₃₆ with DBE 1-3. These formulas most likely represent compound classes with terminal functional groups such as fatty acids, primary alcohols and

aldehydes or esters (Riederer and Muller, 2008). However, the even numbered carbon prevalence diminishes with increasing DBE for all humic acid samples until there is very little preference or none at all.

Formulas with H/C ratios approaching 2 and minimal DBEs have been shown to be created after oxidation of lignin extracts with hydroxyl radical by our group. They are also found here to be part of COO Kendrick mass defect series. Therefore, this region has been expanded from the traditional CRAM region to include formulas with H/C higher than 1.5 and slightly lower than 0.25 O/C proposed for CRAM (Hertkorn et al., 2006). Together with the condensed aromatic and lignin-like molecules, these are of interest when considering humification.

5. CONCLUSIONS

Recent application of ultrahigh resolution ESI-FTICR-MS to the study of humic substances has revealed a commonality among humic acids from varied sources and environmental conditions. In general, we can group the components of humic acids into three predominant types that, based on recent findings of organic matter-hydroxyl radical chemistry, may be inter-related chemically. These molecular types refine the current understanding of humic acids, and may denote a pathway for the process of humification of organic matter.

Lignin-like molecules that plot in the moderate H/C and O/C region of the van Krevelen diagram may represent the first stage of humification by microbial and fungal enzymatic attack. Further reactions of these molecules with free radicals in soils, either concerted with efforts by microorganisms to access more readily available energy sources such as carbohydrates, or fortuitously due to the presence of hydroxyl radicals, can account for the remaining two types of molecules. Condensed aromatics, often containing oxygenated functional groups such as hydroxyl and carboxylic acid and previously assumed only to be present as a result of thermal processes are now recognized as a potential product of lignin humification. Similarly, CCAM and CRAM-like molecules, though not physically isolated have mostly been identified in samples from marine environments and hypothesized to be present as a result of degraded microbial biomass, are also a possible byproduct of condensation and decarboxylation reactions of lignin-like molecules in soils. The prevalence of these two groups of molecules relative to lignin-like molecules in humic acids, and the fact that many are included in COO Kendrick mass defect series also correlates well with the degree of humification.

The abundance of the biopolymer lignin and the presence of these molecular groups in far-reaching environments including the open ocean has potentially significant implications for the cycling of terrestrial organic matter. This work represents renewed evidence for lignin as a potential source of much of the organic matter in humic acids from a variety of environments.

499

ACKNOWLEDGEMENTS

500 This work was funded by DOE SBR grant #DE-ER64567-1031562-0014364. We thank Dr.
501 Tsutomu Ohno for making the data from his 2010 study available to us. We also thank the
502 COSMIC (College of Science Major Instrumentation Cluster) facility at Old Dominion
503 University for allowing us to use their equipment.

504

505

REFERENCES

506 Antony R., Grannas A.M., Willoughby A.S., Sleighter R.L., Thampan M. and Hatcher P.G.
507 (2014) Origin and Sources of Dissolved Organic Matter in Snow on the East Antarctic
508 Ice Sheet. *Environ. Sci. Technol.* **48**, 6151-6159.

509

510 Boerjan W., Ralph J. and Baucher M. (2003) Lignin biosynthesis. *Annu. Rev. Plant Biol.* **54**,
511 519-546.

512

513 Carlile M.J., Watkinson S.C. and Gooday G.W. (2001) *The Fungi*, Second ed. Elsevier Ltd,
514 Hungary.

515

516 Chefetz B., Chen Y., Clapp C.E. and Hatcher P.G. (2000) Characterization of organic matter in
517 soils by thermochemolysis using tetramethylammonium hydroxide (TMAH). *Soil Sci.*
518 *Soc. Am. J.* **64**, 583-589.

519

520 Chen C.-L., Chang H.-M. and Kirk T.K. (1983) Carboxylic Acids Produced Through Oxidative
521 Cleavage of Aromatic Rings During Degradation of Lignin in Spruce Wood by
522 Phanerochaete Chrysosporium. *J. Wood Chem. Technol.* **3**, 35-57.

523

524 Chen H., Abdulla H.A.N., Sanders R.L., Myneni S.C.B., Mopper K. and Hatcher P.G. (2014)
525 Production of Black Carbon-like and Aliphatic Molecules from Terrestrial Dissolved
526 Organic Matter in the Presence of Sunlight and Iron. *Environ. Sci. Technol. Lett.* **1**, 399-
527 404.

528

529 Chen Y. and Schnitzer M. (1978) The surface tension of Aqueous Solutions of Soil Humic
530 Substances. *Soil Science* **125**, 7-15.

531

532 Grinhut T., Hertkorn N., Schmitt-Kopplin P., Hadar Y. and Chen Y. (2011) Mechanisms of
533 Humic Acids Degradation by White Rot Fungi Explored Using ¹H NMR Spectroscopy
534 and FTICR Mass Spectrometry. *Environ. Sci. Technol.* **45**, 2748-2754.

535

- Guggenberger G. (2005) Humification and Mineralization in Soils, in: Varma A., Buscot F. (Eds.), *Microorganisms in Soils: Roles in Genesis and Functions*. Springer Berlin, pp. 85-106.
- Gupta N.S. (2014) *Biopolymers: A molecular paleontology approach*. Springer Netherlands.
- Hartman B.E., Chen H. and Hatcher P.G. (2015) A non-thermogenic source of black carbon in peat and coal. *International Journal of Coal Geology* **144–145**, 15-22.
- Hatcher P.G. (1988) Dipolar-dephasing ¹³C NMR studies of decomposed wood and coalified xylem tissue: evidence for chemical structural changes associated with defunctionalization of lignin structural units during coalification. *Energy & Fuels* **2**, 48-58.
- Hatcher P.G. and Clifford D.J. (1994) Flash Pyrolysis and in-situ methylation of humic acids from soil. *Org. Geochem.* **21**, 1081-1092.
- Hertkorn N., Benner R., Frommberger M., Schmitt-Kopplin P., Witt M., Kaiser K., Kettrup A. and Hedges J.I. (2006) Characterization of a major refractory component of marine dissolved organic matter. *Geochim. Cosmochim. Acta* **70**, 2990-3010.
- Higuchi T. (2004) Microbial degradation of lignin: Role of lignin peroxidase, manganese peroxidase, and laccase. *P Jpn Acad B-Phys* **80**, 204-214.
- Hockaday W.C., Grannas A.M., Kim S. and Hatcher P.G. (2006) Direct molecular evidence for the degradation and mobility of black carbon in soils from ultrahigh-resolution mass spectral analysis of dissolved organic matter from a fire-impacted forest soil. *Org. Geochem.* **37**, 501-510.
- Hur M., Yeo I., Park E., Kim Y.H., Yoo J., Kim E., No M.-h., Koh J. and Kim S. (2010) Combination of Statistical Methods and Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for More Comprehensive, Molecular-Level Interpretations of Petroleum Samples. *Anal. Chem.* **82**, 211-218.
- Ikeya K., Sleighter R.L., Hatcher P.G. and Watanabe A. (2013) Fourier transform ion cyclotron resonance mass spectrometric analysis of the green fraction of soil humic acids. *Rapid Commun. Mass Spectrom.* **27**, 2559-2568.

- Ikeya K., Sleighter R.L., Hatcher P.G. and Watanabe A. (2015) Characterization of the chemical composition of soil humic acids using Fourier transform ion cyclotron resonance mass spectrometry. *Geochim. Cosmochim. Acta* **153**, 169-182.
- Kamga A.W., Behar F. and Hatcher P.G. (2014) Quantitative Analysis of Long Chain Fatty Acids Present in a Type I Kerogen Using Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: Compared with BF₃/MeOH Methylation/GC-FID. *J. Am. Soc. Mass. Spectrom.* **25**, 880-890.
- Kim S., Kramer R.W. and Hatcher P.G. (2003) Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van Krevelen diagram. *Anal. Chem.* **75**, 5336-5344.
- Koch B.P. and Dittmar T. (2006) From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. *Rapid Commun. Mass Spectrom.* **20**, 926-932.
- Koch B.P., Ludwighowski K.-U., Kattner G., Dittmar T. and Witt M. (2008) Advanced characterization of marine dissolved organic matter by combining reversed-phase liquid chromatography and FT-ICR-MS. *Mar. Chem.* **111**, 233-241.
- Koenig A.B., Sleighter R.L., Salmon E. and Hatcher P.G. (2010) NMR Structural Characterization of *Quercus alba* (White Oak) Degraded by the Brown Rot Fungus, *Laetiporus sulphureus*. *J. Wood Chem. Technol.* **30**, 61-85.
- Kögel-Knabner I., Zech W. and Hatcher P.G. (1991) Chemical Structural Studies of Forest Soil Humic Acids: Aromatic Carbon Fraction. *Soil Sci. Soc. Am. J.* **55**, 241-247.
- Kramer R.W., Kujawinski E.B. and Hatcher P.G. (2004) Identification of Black Carbon Derived Structures in a Volcanic Ash Soil Humic Acid by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Environ. Sci. Technol.* **38**, 3387-3395.
- Kramer R.W., Kujawinski E.B., Zang X., Green-Church K.B., Jones R.B., Freitas M.A. and Hatcher P.G. (2001) Studies of the structure of humic substances by electrospray ionization coupled to a quadrupole-time of flight (QQ-TOF) mass spectrometer, in: Ghabbour E.A., Davies G. (Eds.), *Humic Substances: Structures, Models and Functions*. The Royal Society of Chemistry, Cambridge, pp. 95-107.
- Kumada K. (1987) *Chemistry of Soil Organic Matter*. Elsevier, Amsterdam.

- Kurbatov I.M. (1968) *The question of the genesis of peat and its humic acids.*, in: Robertson R.A. (Ed.), Transactions of the 2nd International Peat Congress, Leningrad, HMSO, Edinburgh., pp. 133-137.
- Lam B., Baer A., Alae M., Lefebvre B., Moser A., Williams A. and Simpson A.J. (2007) Major Structural Components in Freshwater Dissolved Organic Matter. *Environ. Sci. Technol.* **41**, 8240-8247.
- Leonowicz A., Cho N.S., Luterek J., Wilkolazka A., Wojtas-Wasilewska M., Matuszewska A., Hofrichter M., Wesenberg D. and Rogalski J. (2001) Fungal laccase: properties and activity on lignin. *J. Basic Microbiol.* **41**, 185-227.
- Mesfioui R., Love N.G., Bronk D.A., Mulholland M.R. and Hatcher P.G. (2012) Reactivity and chemical characterization of effluent organic nitrogen from wastewater treatment plants determined by Fourier transform ion cyclotron resonance mass spectrometry. *Water Res.* **46**, 622-634.
- Nebbioso A. and Piccolo A. (2013) Molecular characterization of dissolved organic matter (DOM): a critical review. *Anal. Bioanal. Chem.* **405**, 109-124.
- Nebbioso A., Piccolo A., Lamshoft M. and Spiteller M. (2014) Molecular characterization of an end-residue of humeomics applied to a soil humic acid. *Rsc Advances* **4**, 23658-23665.
- Ohno T., He Z.Q., Sleighter R.L., Honeycutt C.W. and Hatcher P.G. (2010) Ultrahigh Resolution Mass Spectrometry and Indicator Species Analysis to Identify Marker Components of Soil- and Plant Biomass-Derived Organic Matter Fractions. *Environ. Sci. Technol.* **44**, 8594-8600.
- Olk C., Dancel M.C., Moscoso E., Jimenez R.R. and Dayrit F.M. (2002) Accumulation of lignin residues in organic matter fractions of lowland rice soils; A pyrolysis-GC-MS study. *Soil Science* **167**, 590-606.
- Page S.E., Sander M., Arnold W.A. and McNeill K. (2012) Hydroxyl Radical Formation upon Oxidation of Reduced Humic Acids by Oxygen in the Dark. *Environ. Sci. Technol.* **46**, 1590-1597.
- Page S.E., Kling G.W., Sander M., Harrold K.H., Logan J.R., McNeill K. and Cory R.M. (2013) Dark Formation of Hydroxyl Radical in Arctic Soil and Surface Waters. *Environ. Sci. Technol.* **47**, 12860-12867.

- Reemtsma T. (2009) Determination of molecular formulas of natural organic matter molecules by (ultra-) high-resolution mass spectrometry: Status and needs. *J. Chromatogr. A* **1216**, 3687-3701.
- Riederer M. and Muller C. (2008) *Annual Plant Reviews, Biology of the Plant Cuticle*. Wiley.
- Ruiz-Dueñas F.J. and Martínez Á.T. (2009) Microbial degradation of lignin: how a bulky recalcitrant polymer is efficiently recycled in nature and how we can take advantage of this. *Microbial biotechnology* **2**, 164-177.
- Schulten H.R. and Schnitzer M. (1992) Structural studies on soil humic acids by curie-point pyrolysis-gas chromatography mass-spectrometry *Soil Science* **153**, 205-224.
- Simpson A.J., McNally D.J. and Simpson M.J. (2011) NMR spectroscopy in environmental research: From molecular interactions to global processes. *Prog. Nucl. Magn. Reson. Spectrosc.* **58**, 97-175.
- Singer G.A., Fasching C., Wilhelm L., Niggemann J., Steier P., Dittmar T. and Battin T.J. (2012) Biogeochemically diverse organic matter in Alpine glaciers and its downstream fate. *Nature Geosci* **5**, 710-714.
- Sleighter R.L. and Hatcher P.G. (2007) The application of electrospray ionization coupled to ultrahigh resolution mass spectrometry for the molecular characterization of natural organic matter. *J. Mass Spectrom.* **42**, 559-574.
- Sleighter R.L. and Hatcher P.G. (2008) Molecular characterization of dissolved organic matter (DOM) along a river to ocean transect of the lower Chesapeake Bay by ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Mar. Chem.* **110**, 140-152.
- Sleighter R.L., McKee G.A. and Hatcher P.G. (2009) Direct Fourier transform mass spectral analysis of natural waters with low dissolved organic matter. *Org. Geochem.* **40**, 119-125.
- Sleighter R.L., McKee G.A., Liu Z. and Hatcher P.G. (2008) Naturally present fatty acids as internal calibrants for Fourier transform mass spectra of dissolved organic matter. *Limnol. Oceanogr. Methods* **6**, 246-253.

- Sleighter R.L., Chen H., Wozniak A.S., Willoughby A.S., Caricasole P. and Hatcher P.G. (2012) Establishing a Measure of Reproducibility of Ultrahigh-Resolution Mass Spectra for Complex Mixtures of Natural Organic Matter. *Anal. Chem.* **84**, 9184-9191.
- Sollins P., Swanston C. and Kramer M. (2007) Stabilization and destabilization of soil organic matter—a new focus. *Biogeochemistry* **85**, 1-7.
- Stevenson F.J. (1994) *Humus Chemistry: Genesis, Composition, Reactions*. John Wiley & Sons, Canada.
- Stubbins A., Spencer R.G.M., Chen H.M., Hatcher P.G., Mopper K., Hernes P.J., Mwamba V.L., Mangangu A.M., Wabakanghanzi J.N. and Six J. (2010) Illuminated darkness: Molecular signatures of Congo River dissolved organic matter and its photochemical alteration as revealed by ultrahigh precision mass spectrometry. *Limnol. Oceanogr.* **55**, 1467-1477.
- Sutton R. and Sposito G. (2005) Molecular structure in soil humic substances: The new view. *Environ. Sci. Technol.* **39**, 9009-9015.
- URL: *USGS Federal Fire Occurrence Website*,
[\[http://wildfire.cr.usgs.gov/firehistory/index.html\]](http://wildfire.cr.usgs.gov/firehistory/index.html).
- URL: *USDA Natural Resources Conservation Service Web Soil Survey*,
[\[http://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx\]](http://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx).
- URL: *Official Soil Series Descriptions (OSD) with series extent mapping capabilities*,
<https://soilseries.sc.egov.usda.gov/>.
- van Krevelen D. (1950) Graphical-statistical method for the study of structure and reaction processes of coal *Fuel* **29**, 269-284.
- Waggoner D.C., Chen H., Willoughby A.S. and Hatcher P.G. (2015) Formation of black carbon-like and alicyclic aliphatic compounds by hydroxyl radical initiated degradation of lignin. *Org. Geochem.* **82**, 69-76.
- Watanabe A., Fujimori H., Nagai Y., Miyajima T. and Kuwatsuka S. (1996) Analysis of the green fraction of humic acids. *Eur. J. Soil Sci.* **47**, 197-204.

729 Whiteside G.B. (1950) *Soil Survey of Prince Edward Island*, in: Agriculture C.D.o., Agriculture
730 P.E.I.D.o. (Eds.). Experimental Farms Service.

731

732

733

Figure Captions

Figure 1. Van Krevelen diagram for humic acids extracted from a grassland soil in Valentine, NE. Data were obtained from Ohno et al. (2010). Only formulas containing the atoms C, H, and O are plotted. Blue, yellow, and gray regions define CCAM, lignin-like, and black carbon-like (BC) regions, respectively.

Figure 2. Van Krevelen diagrams for humic acids from a) a deciduous forest/swamp (GA-1) and b) a grassy/open field (GA-2) in Kingsland, GA. CHO only compounds are plotted.

Figure 3. CPMAS and DPMAS spectra of humic acids GA-1 a) CPMAS and c) DPMAS; GA-2 humic acids b) CPMAS and d) DPMAS. Spinning side bands are denoted with a black dot.

Figure 4. Van Krevelen diagrams for humic acids from various soil ecosystems and soil types. a) agricultural/native grassland in Catlin, IL, b) deciduous forest spodosol in the Armadale region of Prince Edward Island, Canada, c) the buried humic layer of a highly humified forest soil in Japan (reproduced from Ikeya et al. (2013) with permission). Only CHO compounds are plotted.

Figure 5. Relative distribution of molecular formulas and formula intensities of calcium humic acids in three identified regions of the van Krevelen diagram (carboxyl containing aliphatic molecules (CCAM), lignin-like, and condensed aromatic). Humic acids are from a variety of soil types and ecosystems at USDA sites across the US (data from Ohno et al. (2010)).

Figure 6. Percentage of formulas in humic acid samples that are part of COO Kendrick mass defect series of 2 or more.

Figure 7. a) Van Krevelen diagram for calcium humic acids from a USDA site in Catlin, IL. b) COO Kendrick mass defect per carbon number for Catlin, IL humic acid molecular formulas.

Figure 8. Percentage of formula that are part of a COO Kendrick mass defect series in each van Krevelen region of humic acid samples.

Figure 9. Van Krevelen diagram for humic acids extracted from a grassland soil in Valentine, NE (blue) with molecular formula in common with photochemically-generated formulas shown by Chen et al. (2014) overlaid in red. Only the CHO compounds are plotted.

Figure 10. Intensity-weighted DBE vs. carbon number plot for the Armadale humic acid. Bubble size represents the relative individual peak intensities.

763 Table 1

764

765 Properties of humic acid parent soils.

766	Sample Name	Location	Land use	Texture Class	Drainage
767	Valentine ^a	Cherry County, NE	Rangeland	Sand	Good
768	GA-1 ^b	Kingsland, GA	Forest Swamp	Loam/sand	Poor
769	GA-2 ^b	Kingsland, GA	Open Field	Sand	Good
770	Catlin ^a	Ogle, IL	Agriculture	Silt/clay/loam	Moderate
771	Armada ^c	Prince Edward Island,	Forest	Sandy loam/clay loam	Poor
772		Canada			
773	Tsubame ^d	Niigata Prefecture,	Sub Alpine Forest	N/A	Good
774		Japan			

775 ^aUSDA Official Soil Series Descriptions (OSD) website

776 ^bUSDA Natural Resources Conservation Service Web Soil Survey

777 ^cWhiteside (1950)

778 ^dKumada (1987), Watanabe et al. (1996)

779 Table 2

780

781 Assigned CHO formulas and summed peak magnitudes for Van Krevelen Regions

782 Sample 783 Name	Total No. 783 Formulas 784 Assigned	No. CHO 783 Formulas 784 Assigned	782 % CHO Formulas (% CHO Peak Magnitude)			782 % Formulas 783 in COO series 784 of 2 or more
			CCAM	Lignin-like	Condensed Aromatic	
785 Valentine	881	573	41 (78.3)	14 (3.2)	47 (16.9)	75
786 GA-1	925	419	67 (64.8)	36 (32.5)	9 (10.3)	55
787 GA-2	1206	633	50 (42.6)	42 (39.7)	25 (32)	70
788 Catlin	1064	775	71 (90.2)	3 (0.3)	27 (9.5)	81
789 Armadale	760	384	74 (76.4)	12 (8.6)	16 (14.6)	53

790

791 Table 3

792 DPMAS Integral Areas (% distribution of carbon signal) and ratio of Aryl-O to Aromatic Carbon Signal

793

Chemical Shift (ppm)								
Sample	Aliphatic (0-45)	Methoxy (45-62)	Alkyl-O (62-120)	Aromatic (120-140)	Aryl-O (140-160)	Carboxyl (160-185)	Carbonyl (185-210)	Ratio Aryl-O: Aromatic
GA-1	21.1	7.8	27.0	13.4	12.3	14.2	4.3	0.91
GA-2	15.0	6.9	31.2	20.0	11.8	13.1	2.0	0.59

798

799

800 Table 4

801 CPMAS Integral Areas (% distribution of carbon signal) and Fraction of Non-Protonated Carbon in GA-1 and GA-2 Humic Acids

802	803 <u>Chemical Shift (ppm)</u>		804 Aliphatic Region Normalization Factor ^a	805 Normalized Area of CPMAS Aromatic Region	806 % Increase in Area of Aromatic Region from CPMAS
	807 Aliphatic (0-45)	808 Aromatic (120-140)			
809	GA-1	35.2	10.8	1.67	22.4
810	GA-2	25.3	17.8	1.69	33.7

811 ^aRatio of aliphatic region peak areas integrated in CPMAS to that integrated in DPMAS (Table 3).

812

813

814

Figure 1

Figure 4

Figure 5

Figure 6

% Formulas in COO Series

90%
80%
70%
60%
50%
40%
30%
20%
10%
0%

Valentine

Catlin

Armada

GA-1

GA-2

Figure 8

Figure 9

Figure 10