Identification of ecosystem-specific markers in terrestrial dissolved organic matter (DOM) by Orbitrap mass spectrometry

Dissertation

zur Erlangung des akademischen Grades doctor rerum naturalium

(Dr. rer. nat.)

vorgelegt dem Rat der Chemisch-Geowissenschaftliche Fakultät der

Friedrich-Schiller-Universität Jena

von M.Sc. Marine Umweltwissenschaften Carsten Simon

geboren am 07.06.1988 in Potsdam

Gutachter:

1. ....

2. ....

Tag der Verteidigung: ....

Danksagung

„Es ist ganz wahr, was die Philosophie sagt, dass das Leben rückwärts verstanden werden muss.

Aber darüber vergisst man den andern Satz, dass vorwärts gelebt werden muss."

Søren Kierkegaard, Tagebücher 1834 - 1855

Ich möchte mich an dieser Stelle bei all jenen bedanken, die mich während der gesamten Doktorandenzeit begleitet, unterstützt und abgelenkt haben, und auch bei einigen, die mich auf diese Zeit unbewusst vorbereitet haben.
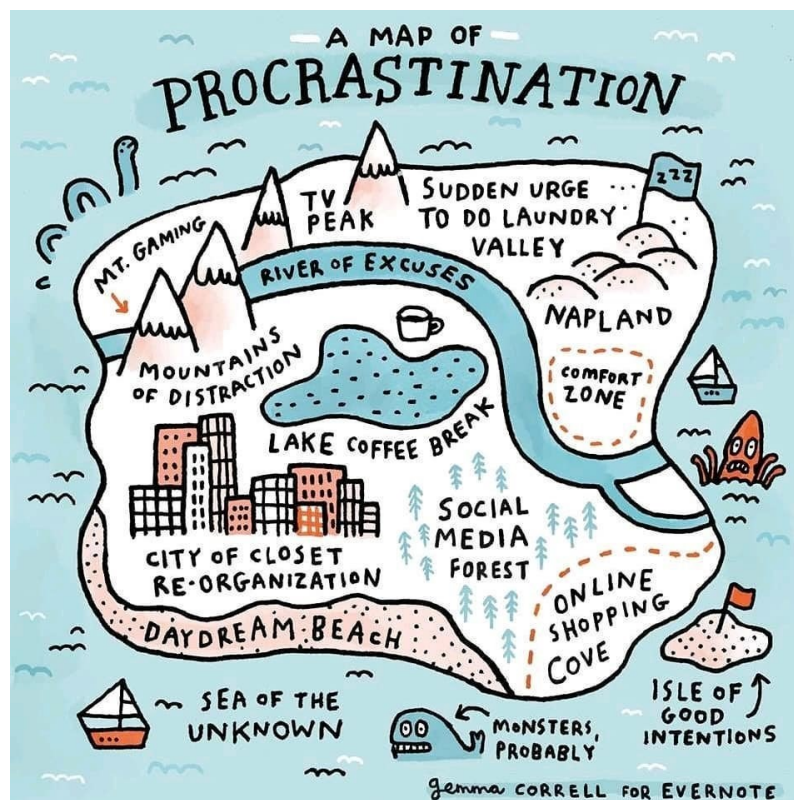
„A map of procrastination" von Gemma Correll für Evernote (https://www.gemmacorrell.com/).

Zuallererst gilt mein Dank apl. Prof. Dr. Gerd Gleixner am Max-Planck-Institut für Biogeochemie und Prof. Dr. Georg Pohnert an der Friedrich-Schiller-Universität für die Betreuung meiner Doktorarbeit.

Ich danke Dir, Gerd, für die Möglichkeit, mit Dir am Max-Planck-Institut die Promotion anzugehen, und auch für das damit verbundene Vertrauen. Du hast mir ein wahrhaft spannendes

Thema vorgelegt und mir alle Möglichkeiten der Welt gegeben mich darin auszutoben. Manchmal musste ich feststellen, dass ich vor lauter Möglichkeiten in Sackgassen gelaufen war, oder den Wald vor lauter Bäumen nicht mehr gesehen habe. Du hast mir in solchen Momenten oft den Weg gewiesen, und mich so wieder auf die richtige Bahn gelenkt. Du hast mir tatsächlich das Gefühl vermittelt, ein „Doktorsohn" zu sein.

Dir, Georg, möchte ich für die bereitwillige Übernahme des Erstgutachtens danken. Du hast mir in den Sitzungen meines Beratungskomittees wahrhaft zur Seite gestanden und maßgeblich dazu beigetragen, dass mein Forschungsaufenthalt mich in völlig neue Denkwelten eingeführt hat.

Prof. Dr. Thorsten Dittmar möchte ich für die Teilnahme an meinem Beratungskommittee danken; das Wissen, dass ich während der Masterarbeit unter deiner Betreuung sammeln konnte, hat mich ein gutes Stück durch die Doktorarbeit getragen, methodisch (Massenspektrometrie) und auch thematisch (DOM, Brasilien). Bei den neuen Herausforderungen hast du mich dann aus der Ferne mit aller Kraft unterstützt.

Aufenthalte in Japan, Brasilien und den USA waren die Highlights meiner Promotionszeit. Mein tiefer Dank gilt daher Prof Dr. Pieter Dorrestein, Dr. Alberto Quesada, Dr. Tania Pimentel und Dr. Savio Ferreira, die mir Aufenthalte in ihren Laboren und Feldstationen ermöglicht haben. Ich bedanke mich dahingehend für die großzügige Unterstützung durch meine Graduiertenschule, die IMPRS-gBGC und das vom Bundesministerium für Bildung und Forschung (BMBF) und weiteren brasilianischen Projektträgern geförderte deutsch-brasilianische Projekt ATTO (Amazon Tall Tower Observatory). Ich möchte mich an dieser Stelle auch für das Reisestipendium der International Humic Substances Society bedanken, das mir die Teilnahme an der 16. Internationalen Konferenz der IHSS in Kanazawa, Japan, ermöglicht hat. Sam Jones, Thomas Behrendt, Stefan Wolff und Fernanda da Luz waren (und sind) wichtige Stützen während der insgesamt zweimonatigen Arbeiten in Brasilien. Daniel Petras und seine Kollegen waren unersetzlich in San Diego. Ich hoffe, dass die inspirierende Zusammenarbeit mit Euch bald Früchte trägt!

Mein besonderer Dank gilt Dr. Vanessa-Nina Roth, die mich während ihrer Zeit am Max-Planck-Institut tatkräftig unterstützt hat und in vielerlei Hinsicht die Grundlage für meine Doktorarbeit gelegt hat. Deine Arbeitsmoral, deine Forderung und Förderung meiner Person

haben mir wichtige Leitplanken gesetzt. Dein Weggang war ein Einschnitt, aber er hat mir den nötigen Ansporn gegeben, um selbstständig weiter zu gehen.

Ich danke allen Arbeitsgruppenmitgliedern und Departmentkollegen, die mich über die Jahre begleitet und unterstützt haben. Insbesondere gilt mein Dank Markus Lange, der immer ein offenes Ohr für Fragen hatte. Ich danke Dir für den Austausch und die wahrhaft fruchtbare Teamarbeit. Ich hoffe, dass das so bleibt! Weiterhin danke ich meinen gesprächigen Bürokollegen: Somak, Jianbei, Perla, und Simon. Die Arbeitsatmosphäre mit Euch war immer stimulierend für mich, beruflich wie persönlich. Ihr wart und seid mir alltägliche Vorbilder. Uta, Steffen, Roman, Franzi, Jeetendra, Ronny, Philipp, Sebastian, Julia, Stefan, Su, Kasun, Natalie und Yan saßen mit mir während meiner Doktorarbeit in einem Boot und haben kräftig mitgerudert. Ich hoffe, dass auch ich Ihnen eine Stütze war, in den schwierigen wie den schönen Momenten. John, Gan, Jenia, Holger, Chirag, Elisa, Marcus, Jeff, Antonios, Shane, Manuel und Ingo haben mir öfters über die Mittagspausen, den Kaffee auf dem Flur, oder das Feierabendbier hinweggeholfen. Gemeinsam haben wir dabei unser Leid geteilt und die Freude über das Erreichte verdoppelt. Sebastian, Julia und Miriam möchte ich für ihren engagierten Einsatz danken und für die Gelegenheit Betreuungserfahrung zu sammeln.

Meine tägliche Arbeit am Institut wäre nicht möglich gewesen ohne den Einsatz vieler anderer Menschen. Ich möchte mich bedanken bei Iris Kuhlmann, bei Heiko Moossen, Ines Hilke, Axel Steinhof und ihren Mitarbeitern. Sie haben Messungen ermöglicht, die sonst nicht möglich gewesen wären, oftmals unter persönlichem Mehraufwand. Bertram Smolny und seinen Mitarbeitern möchte ich für ihre oftmals unsichtbare tägliche Hilfe bei Computer-, Drucker- und Internetproblemen danken. René Schwalbe und seinen Mitarbeitern möchte ich für ihre stetige Arbeit im Haus danken. Die Arbeitsbedingungen in den Laboren und den Büros waren dank Euch außergewöhnlich. Frank Voigt und Bernd Schlöffel möchte ich besonders für ihren engagierten Einsatz bei der Konstruktion von Labor- und Feldequipment danken. Auch den Mitarbeitern von Olaf Kolle in der Abteilung Freiland will ich für die bereitwillige Herausgabe von Ratschlägen und Feldequipment danken. Unsere Arbeit in Brasilien wäre ohne die Zuarbeit von Mechanik und Freiland so nicht möglich gewesen. Der Verwaltung, Bibliothek und Steffi und John im IMPRS-Büro möchte ich an diese Stelle für ihre organisatorische Leitung im Hintergrund des alltäglichen Wissenschafts-Geschäfts danken.

Auch meinen vielen Lehrern sei gedankt für ihre Beihilfe zur Promotion. Meiner ersten „Lehrerin", Oma Dorchen, danke ich dafür, dass Du mein Interesse für die Erde von

Dinosauriern auch auf die etwas unscheinbareren Dinge, wie Steine, Wolken, und den Boden auf dem wir stehen gelenkt hast. Den Grundschullehrerinnen Frau Große und Frau Kerber danke ich für die frühe Förderung, und auch den Gymnasiallehrern Fr. Freydank, Fr. Naruhn, Hr. Dr. Fischer und Hr. Hering sei ihr Einsatz rückwirkend noch einmal gedankt. Ihr Unterricht hat über das reine Fachwissen hinausgewirkt. Den Lehrenden in Jena und Oldenburg danke ich für die Übernahme des Staffelstabs „Lehre": Beate Michalzik, Kai-Uwe Totsche, Lothar Viereck und Kirsten Küsel haben so meine Interessen in Jena besonders gelenkt. In Oldenburg waren es dann Barbara Scholz-Böttcher, Bernhard Schnetger, Jürgen Köster, Cora Kohlmeier und Verona Vandieken, die unter hohem persönlichem Aufwand und durch interessante und fordernde Lehre mein Studium bereichert haben.

Das Leben besteht nicht nur aus Arbeit, auch wenn es oft so schien in den letzten Jahren. Ich ziehe meinen Hut vor Steffen, Philipp, Yann und Lena, Doro, Michelle, Frieder, Jomar, Thomas, Nina, Jens, Tizian, Sven, Leo, Anna, Lilly, Robert, Anne Müller, Anne Weiß, Daniel und Nina, und ihren Freunden. Ohne Euch wäre ich in meiner Arbeit untergegangen und hätte vergessen zu leben und mich zu erholen. Dadurch ist mir oft erst aufgefallen, wie unbeschwert ich es eigentlich in dieser Zeit hatte, und wie schwer ich es mir oft gemacht habe.

Meiner Familie und der Familie Gransee und Jurack möchte ich danken für den Rückhalt und das Vertrauen und die offenen Arme, die ihr mir entgegengebracht habt. Gleichzeitig möchte ich mich entschuldigen für meine weitestgehende Abwesenheit über die letzten Jahre hinweg, und die verstrichenen Gelegenheiten, das Leben etwas öfter gemeinsam zu genießen. Meinen Eltern, Maren und Jörn, danke ich für die guten „Startbedingungen", die ihr mir geschaffen habt. Dafür, dass ihr mich auch mal zum Lesen und Rechnen lernen gezwungen habt. Dafür, dass ihr mir beide Welten – sowohl das analytische Denken als auch das kreative Schaffen – nahegebracht habt, und den aus mir gemacht habt, der ich bin. Ich bin stolz auf Euch!

Zuguterletzt bleibt der wichtigste Teil, meine bessere Hälfte. Liebe Aline, niemand hat so sehr unter dieser Doktorarbeit gelitten wie du. Danke das für mich da warst, und entschuldige, dass ich es so oft nicht für dich war. Wir haben wahrlich alle Aufs uns Abs erlebt. In diesem Moment, wo ich diese Zeilen schreibe, bereitest du dich angstvoll auf die Verteidigung deiner Masterarbeit vor. Ich drücke Dir die Daumen, ich weiß um deine Kompetenz. Eigentlich kann doch nichts schief gehen –

Ich bin gespannt auf unsere *gemeinsame* Zukunft!

Table of Contents

Figures

Tables

Abbreviations

μm	Micrometer, 10^{-9} m
AImod	Modified aromaticity index
AMS	Accelerator mass spectrometry
API	Atmospheric pressure ionization
a.s.l.	Above sea level
a.u.	Arbitrary units
BC	Polycyclic, condensed aromates, such as "Black Carbon"
C18	Octadecyl-bonded silica
CID	Collision-induced dissociation
CHO	Formulae classified according to heteroatom content (no N, S, P)
CHOS	Sulfur-containing formulae
CHNO	Nitrogen-containing formulae
CRAM	Carboxyl-rich alicyclic molecules
CZ	Critical Zone
Da	Dalton
DBE	Double bond equivalent
DBE/C	Carbon-normalized DBE
DBE-O	Oxygen-corrected DBE
DNP	Database of Natural Products
DOC	Dissolved organic carbon
DOM	Dissolved organic matter
DON	Dissolved Organic Nitrogen
EC	Electrical conductivity
EE	Extraction Efficiency
EEM	Excitation emission matrix (fluorescence spectroscopy)
e.g.	exempli gratia (for example)
ESI	Electrospray ionization
etc.	et cetera (and so forth)
F14C	Fraction Modern
FDOM	Fluorescent dissolved organic matter
FT-ICR MS	Fourier Transform Ion Cyclotron Resonance Mass Spectrometry
FT-ICR MS	Fourier transform ion cyclotron resonance mass spectrometry
FTMS	Fourier transform mass spectrometry (ultrahigh resolution mass spectrometry)
FWHM	Full width half maximum
GC	Gas chromatography
H/C	Atomic ratio of hydrogen to carbon in a molecular formula
HCD	Higher charge collisional dissociation
HMW	High molecular weight
HU	Highly unsaturated compounds
hw	(as a prefix to BC/ PP) higher molecular weight; ≥ 15 C atoms in formula
i.e.	id est (in other words)

IHSS	International Humic Substances Society
INPA	National Institute for Amazonian Research
IPIM; IPIMs	Isobaric precursor ion mixture(s)
IR	Infrared
LC	Liquid chromatography
LMW	Low molecular weight
lw	(as a prefix to BC/ PP) very low molecular weight; <15 C atoms in formula
MD/ MDs	Mass difference(s)
mDa	Millidalton
MDLT	Material derived from linear terpenoids
MQ	Ultrapure water (deionized water, DI, MilliQ)
ms	Millisecond
MS	Mass spectrometry
MS²	Tandem mass spectrometry, fragmentation mass spectrometry
m/z	Mass-to-charge ratio
[M-H]⁻	"Mother ion"; negatively charged precursor ion; loss of proton
MW	Molecular weight
NMDS	Non-metric multidimensional scaling
NMR	Nuclear magnetic resonance spectroscopy
NOSC	Nominal oxidation state of carbon
NPOC	Non-purgeable organic carbon
O/C	Atomic ratio of oxygen to carbon in a molecular formula
op	(as a prefix to PP, HU & UA) formula poor in oxygen; $O/C \leq 0.5$
or	(as a prefix to PP, HU & UA) formula rich in oxygen; $O/C > 0.5$
PARAFAC	Parallel factor analysis
PC	Polycarbonate
PCA	Principal component analysis
PCoA	Principal coordinates analysis (Metric multidimensional scaling)
PEP	Unsaturated, O- and N-containing compound, such as peptides
pH	-log [H+]
POM	Particulate organic matter
POC	Particulate organic carbon
PP	Polyphenols
ppm	parts per million, mg kg-1
PPL	Modified styrene divinylbenzene polymer
PTFE	Polytetrafluoroethylene
SD	Standard deviation
SFA	Saturated, O-containing compound, such as fatty acids
SID	Source-induced dissociation
s.l.	sensu latu (in the wider sense)
S/N	Signal-to-noise ratio
SOM	Soil organic matter
SOC	Soil organic carbon
SPE	Solid phase extraction

SPE-DOC	Solid-phase-extractable DOC
SRFA	Suwannee river fulvic acid
SRNOM	Suwannee river natural organic matter
s.s.	sensu stricto (in the narrow sense)
SUG	Very high O content, such as sugars
TOC	Total organic carbon
UA	Unsaturated aliphatic compounds
UV	Ultraviolet
VK	Van Krevelen
VOC	Volatile organic carbon
WEOM	Water-extractable organic matter
XAD	Nonionic methyl methacrylate polymer

1   Introduction

1.1   Organic carbon in terrestrial systems

1.1.1   Carbon, life and the Critical Zone

Life on earth is based on the biogeochemistry of the atoms carbon (C), hydrogen (H), oxygen (O), nitrogen (N) and sulfur (S). Other forms of atomic constituents important to life, such as phosphorus (P) or metals (e.g., Mg, K, Ca, Fe, Zn, Mo, Cu, Mn, Ni) depend largely on boundary conditions that are set by the versatility and reactivity of the CHONS system, and the prevalence of aqueous solutions which determine the milieu for reactions (Kleidon, 2016; Langmuir and Broecker, 2012). Soils provide a major interface that mediates ecological interactions involving the water and carbon cycles. For example, plant water uptake relies on water retention, and nutrient availability relies on remineralization of dead biomass as well as retention of nutrients. This way, soils also provide replenishment and decontamination of water resources, sustain plant growth (450 Gt-C globally), and sequester reduced carbon in the form of soil organic matter (SOM, 3000 Gt-C in upper 2 m globally, Figure 1-1; Addiscott, 2010; Bar-On et al., 2018; Kästner and Miltner, 2018; Lavelle et al., 2016; Sanderman et al., 2017) and are thus part of the "Critical Zone" (Ashley, 1998): The Critical Zone is "recognized as a location of complex biogeochemical and physical processes that supports the terrestrial biosphere" and spans from the "outer extent of vegetation down to the lower limits of groundwater" (Brantley et al., 2007; Richardson, 2017). An important aspect of this functioning is the decomposition or remineralization of high-molecular weight and non-soluble organic remains. This process leads not only to the formation of new biomass and more stable forms of soil organic matter (SOM), but also to losses of carbon: Remineralization generates large amounts of gaseous products (volatile organic carbon, VOC) mainly in the form of CO_2 (Hursh et al., 2017). It also produces oxidized lower-molecular weight products that are more soluble than their precursors (dissolved organic carbon, DOC) and escape to deeper soil, groundwater, and aquatic systems such as lakes, rivers, and wetlands (Addiscott, 2010; Kaiser and Kalbitz, 2012).
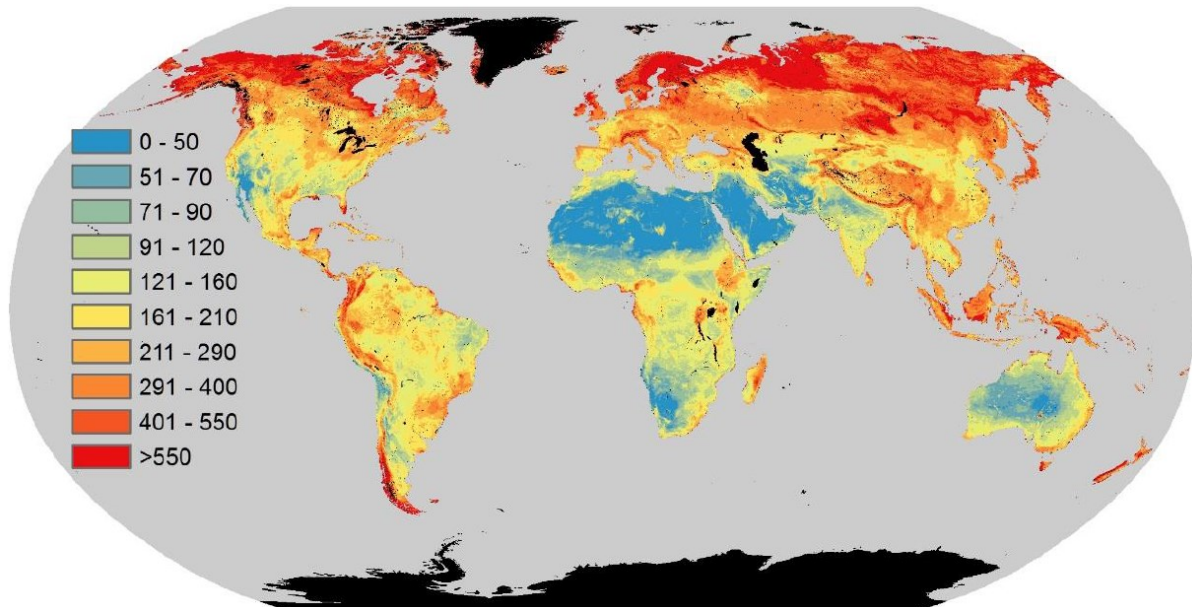
Figure 1-1. Soil organic carbon (SOC) stocks in 0-2 m depth. Estimates given in Mg C ha-1 (Sanderman et al., 2017).

1.1.2 Dissolved organic carbon: Towards molecular biogeochemistry

The fraction of organic molecules of a water sample that pass a filter is operationally defined as the "dissolved" fraction (Filella, 2014; Zsolnay, 2003). The pore size of the applied filter – usually 0.45μm – may however vary from study to study (Dittmar and Stubbins, 2014; Perdue and Ritchie, 2014). Dissolved organic carbon (DOC) represents the major organic matter flux from soils to the aquatic continuum, i.e., rivers, lakes, wetlands, and finally the ocean (Alvarez-Cobelas et al., 2012; Webb et al., 2018). Leaching fluxes from soils and DOC export by rivers have been estimated by up to 10 g-C m^{-2} yr^{-1} at temperate sites with a large amount of DOC retained within soil ($\sim$ 30 g-C m^{-2} yr^{-1}; Borken et al., 2011; Kindler et al., 2010; Michalzik et al., 2001). These numbers show 1) the major role of respiration (on average, 1400 g-C m^{-2} yr^{-1}) for the continuous production of DOC (Hursh et al., 2017; Lee et al., 2018; Malik and Gleixner, 2013) and 2) the high potential for organic matter retention and recycling in deeper layers of soil (Kaiser and Kalbitz, 2012). Moreover, DOC release from soils and watersheds is often described as transport-limited: under most natural conditions, DOC production is only limited by the amount of water flowing through soils (Ledesma et al., 2015; Lee et al., 2018; Zarnetske et al., 2018). The flux of terrestrial DOC is thus largely controlled by the activity of the decomposers, the size of soil organic carbon stocks and the volume and rate of water passage. The manifold effects controlling SOC and DOC destabilization (Figure 1-2) cannot be disentangled by analyzing single properties, such as DOC concentration, radiocarbon

content, stable isotope composition, or optical properties. Recent research suggests that DOC is an ultracomplex mixture of thousands to millions of different structures and better described as DOM (dissolved organic matter; Zark et al., 2017; Zark and Dittmar, 2018). In contrast to the acronym DOC which refers only to the carbon content of organic molecules in solution, the term "dissolved organic matter" (DOM) encompasses the full elemental and structural diversity of these molecules. Following the first description by Piccolo (Piccolo, 2001), DOM mixtures in aqueous solution are now seen as complex, self-assembling and supramolecular associations (Wells and Stretz, 2019). Soluble decomposition products released from soils can be used as tracers of ecosystem functioning: DOM represents a "snapshot of ecosystem activities" (Roth et al., 2014) that contains molecular-level information able to reveal imprints of single sources or processes. However, our knowledge about these potential biomarkers and their response remains scarce (Brown et al., 2016; Hawkes et al., 2019; Malik et al., 2016; Roth et al., 2014). For example, the terrestrial DOC flux "could completely replenish the marine DOC pool within its apparent residence time in the oceans" (Dittmar and Stubbins, 2014). However, the dilute amounts of classical vascular plant biomarkers in marine DOM questions this simple carbon balance (Dittmar and Stubbins, 2014). Much similar, the finding that markers such as lignin show relatively fast turnover in soil (< 50 yrs) has questioned established theories of SOM formation and stability (Austin et al., 2016; Cotrufo et al., 2015; Gleixner, 2013; Marschner et al., 2008; Preston et al., 2009). Although inputs, recycling and transport emerge as the three key factors controlling DOM release from soils and SOM, novel molecular insight is thus needed to properly reveal ecosystem-scale processes, links and responses to environmental change (Kellerman et al., 2018, 2015; Ward et al., 2017).
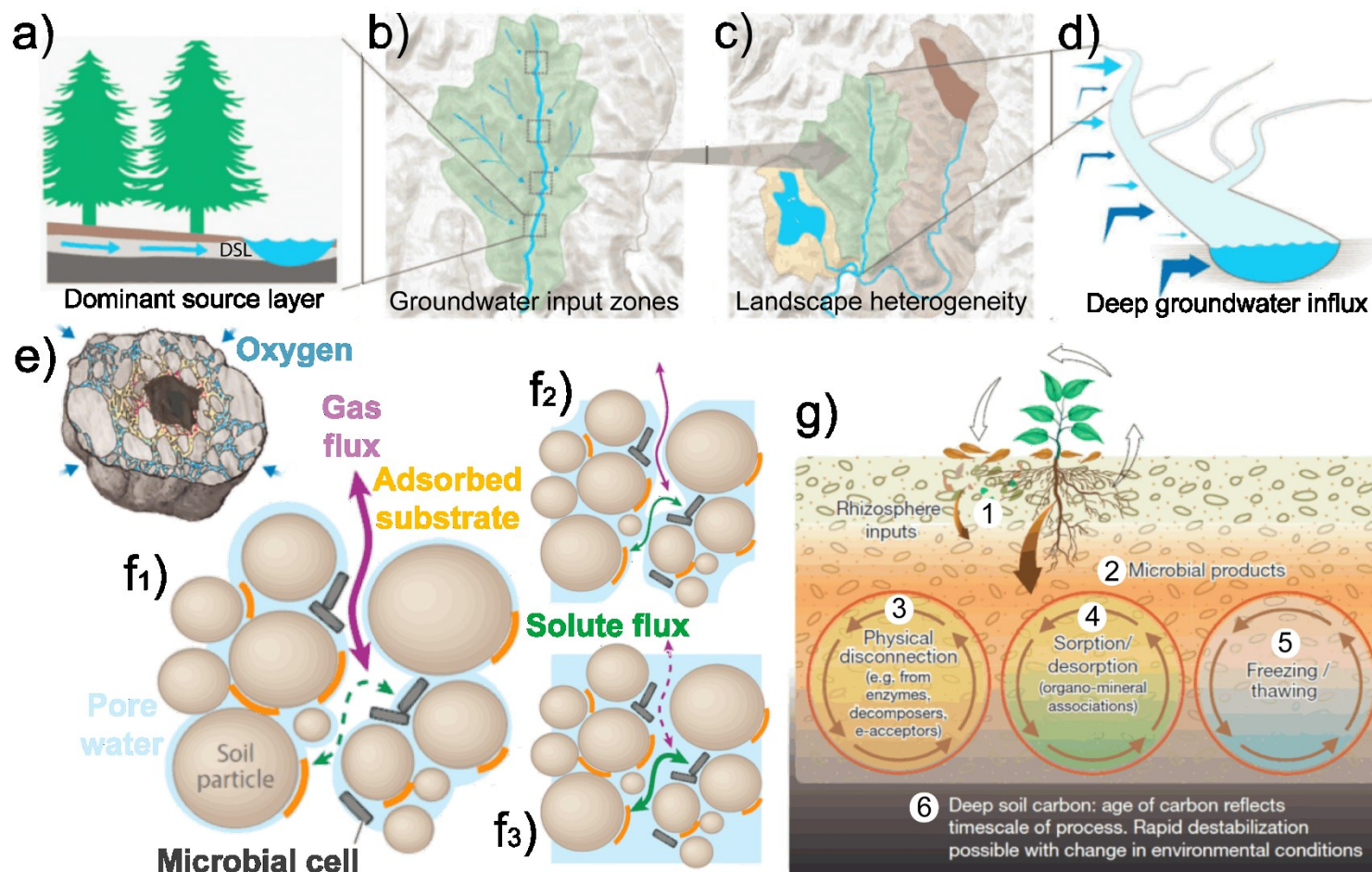
Figure 1-2. Controls on DOM composition, as seen from the catchment (a – d) and soil scale (e – g). Panels a – d) modified from Laudon and Sponseller (2018). e) Soil aggregate section. Oxygen depletion gradient visualized by color. Black object in center is particulate organic matter. Modified from Borer et al. (2018). f1-3) Drying-rewetting cycles in porous media and its influence on gas and solute fluxes. 1) dry, non-saturated conditions, 2) rewetting event, 3) saturated conditions. Modified from Schimel (2018). g) Controls at the soil profile scale. 1) Plant roots and rhizosphere soil inputs, 2) Decomposer community (i.e., fungi, bacteria, invertebrates) and their remains; 3) Heterogeneity and disconnection (i.e., stochastic control); 4) mineral-association, 5) temperature control on decomposition; 6) long-term sequestration. Modified from Schmidt et al. (2011). Note the discrepancy between the "catchment view" of soil (panel a, "dominant source layer") common in hydrology and the fine-scale models developed by soil scientists.
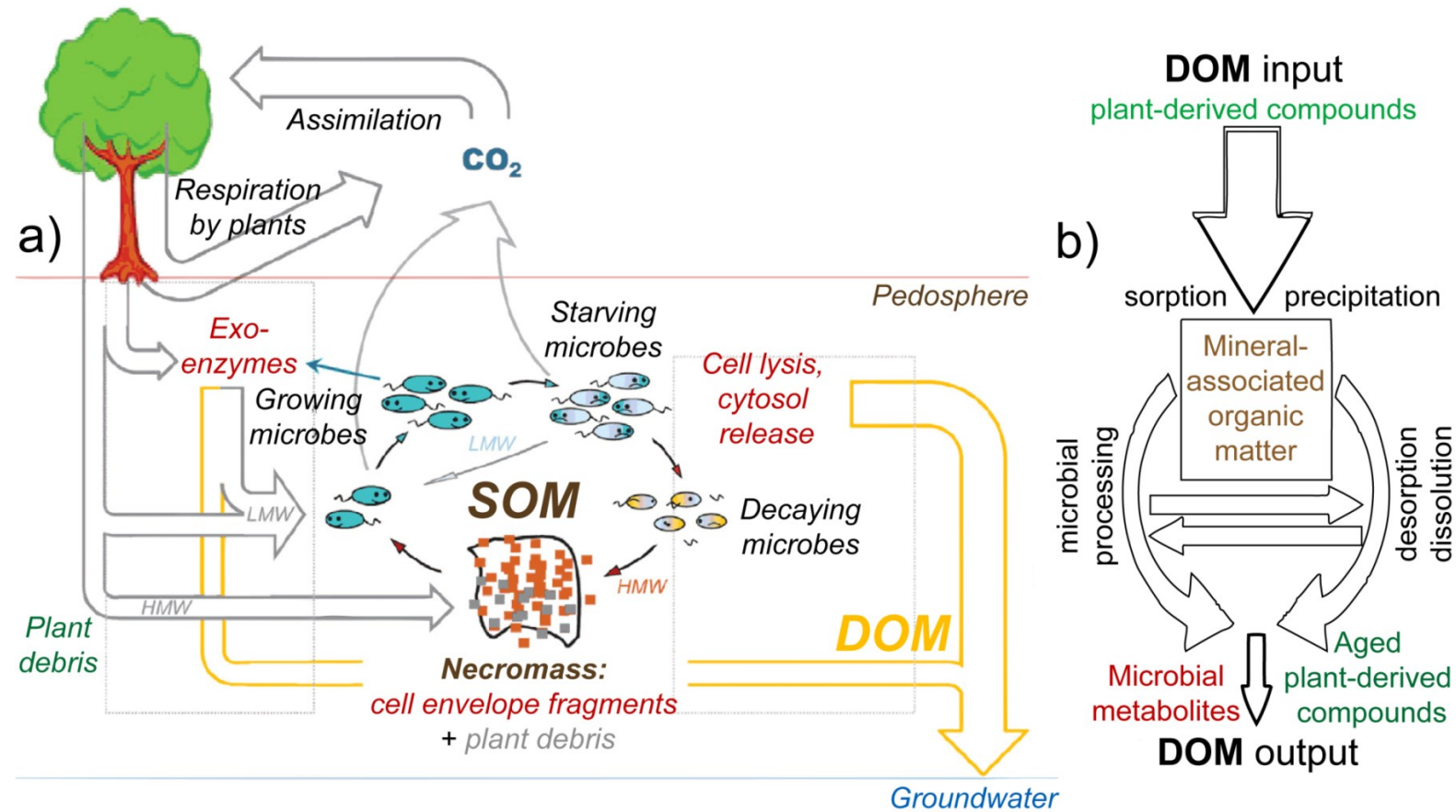
Figure 1-3. The link between SOM and DOM dynamics. **a)** Schematic overview of the microbial carbon pump, taken from Kästner and Miltner (2018): HMW, high molecular weight; LMW, low molecular weight; DOM, dissolved organic matter; SOM, soil organic matter; dotted line, non-ordered extracellular bio(geo)chemistry. Non-soluble plant products (mostly polymers, HMW) are degraded to LMW molecules by exo-enzymes. Part of the material is dissolved and contributes directly to the DOM pool, which may serve as a substrate along its flow path (see panel b). HMW organic matter can be sorbed to particles, allowing some degree of protection from further degradation and contributes to SOM formation. LMW organic matter serves as a substrate for microbes, which subsequently grow, die and lyse. Their necromass forms new SOM, which serves as a substrate for new generations of microbes ("microbial loop"). During decay, soluble products such as in cytosol will contribute to the DOM pool. Microbial cell fragments sorb to soil minerals, and contribute to SOM formation. DOM leaves the system in draining waters (groundwater). **b)** Cascade model proposed by Kaiser & Kalbitz (2012). The model states that reactive transport (e.g., sorption and subsequent reactions) explains lower DOC concentrations in deeper soil. According to the model, this basic processing step repeats along the depth profile ("cascade"; "cycling downwards"), leading to changes in DOM composition.

1.1.3 Sampling of terrestrial DOM

Amount and properties of DOM are affected by the sampling and isolation technique used (Filella, 2014; Zsolnay, 2003). Extraction protocols involving destructive sampling (i.e., homogenization of excavated soil samples, or centrifugation) or the extraction with pH-adjusted (acidic/ alkaline) or heated aqueous solutions usually dissolve fractions of organic matter that are not readily soluble under "natural" conditions, and thus are thought to cause artifacts (Guigue et al., 2014; Lehmann and Kleber, 2015). Under unsaturated conditions, DOC yields depend mainly on soil moisture, and DOM composition indicates significant shifts with increasing matric potential, i.e., when evacuating soil water from ever smaller pore domains that bind water more strongly (Bailey et al., 2017; Zsolnay, 2003). In contrast to non-saturated conditions, groundwater is defined as a water body that resides under saturated conditions, i.e., under fully water-filled pore space, and under zero tension (i.e., flow not affected by the matric potential; Blume et al., 2016). Pore connectivity is thus a similarly important driver of DOM composition (Peyton Smith et al., 2017), and has also been described as a driver of DOM characteristics at the watershed scale (Creed et al., 2003; Fröberg et al., 2006; Hagedorn et al., 2000; Pacific et al., 2010; Seibert et al., 2009; Tunaley et al., 2016).

1.2 DOM: Sources, recycling and transport

Soils are a main source of DOM that may enter aquatic systems, but our understanding of what determines DOM production and turnover in both realms remains vague on the molecular level (Kellerman et al., 2015; Schmidt et al., 2011; Ward et al., 2017). Organic matter properties are controlled by temporal and spatial dynamics of inputs (substrates), recycling (transformation), and transport (outputs; Figure 1-3). This section mainly focuses on soil carbon dynamics (SOM/ DOM); implications for aquatic systems are discussed when relevant within chapters 3 – 6.

1.2.1 Inputs and formation of soil organic matter

DOM stands in constant exchange with SOM (Kaiser and Kalbitz, 2012; Leinemann et al., 2018). Formation of SOM is thus a crucial process that needs to be understood. Potential inputs to soil organic matter come from 1) plant litter (shoots and leaves), 2) root litter, 3) root exudates, 4) faunal biomass and excrements, and 5) microbial biomass (Figure 1-2; Figure 1-3). Subsequently, organic matter input is highly heterogeneous: Highest in hotspots (bio-pores

from roots or fauna, rhizosphere, detritusphere, aggregate surfaces), and lowest in bulk mineral soil (Guggenberger and Kaiser, 2003; Heitkötter and Marschner, 2018; Kuzyakov and Blagodatskaya, 2015; Schimel and Schaeffer, 2012). The DOM pool receives inputs from all of these sources, but the exact source contribution to SOM and DOM formation is a matter of debate (Gross and Harrison, 2019; Sokol et al., 2019). Likewise, the role of DOM for SOM buildup in deeper layers is unclear and shows conflicting results (Bird and Torn, 2006; Cotrufo et al., 2015; Lee et al., 2018; Quan et al., 2018; Scheibe and Gleixner, 2014).

Organic substances found in soils have long been regarded as novel "humic" structures due to their resistance to classical techniques of isolation and identification (Kleber and Lehmann, 2019; Lehmann and Kleber, 2015). NMR data however suggests that they are likely complex mixtures of decomposition products of known compound classes (Masoom et al., 2016; Simpson et al., 2007) such as protein, starch, cellulose, lignin, tannin and other polyphenols, lipids, cutin, suberin, and other natural products of soil microbes and fauna, such as chitin, melanin, or peptidoclucan (Kögel-Knabner 2002). Many lines of evidence point now to an active role ("bottleneck") of the decomposer community in SOM production (Figure 1-3a): Conceptual models ("microbial carbon pump"; Figure 1-3a) originally adopted from oceanography propose that although small in living biomass (< 5 % of SOM), microbes could contribute significantly if their remains would build up to form SOM (Gleixner, 2013; Kästner and Miltner, 2018; Liang et al., 2017). Lab and field data support this hypothesis (Barré et al., 2018; Lange et al., 2015; Lutfalla et al., 2019; Ma et al., 2018; Masoom et al., 2016; Miltner et al., 2012; Simpson et al., 2007; Woche et al., 2017). The classical view of major SOM formation through decomposed plant material thus needs revision (Lehmann and Kleber, 2015; Woolf and Lehmann, 2019). Microscopic techniques have revealed that recognizable remains of microbes and fungi are found widespread in soil microenvironments with clear hotspots related to smaller-sized particles/ pores and rough surfaces, which in part also explain hydrophobic properties of soil (Figure 1-4; Juyal et al., 2018; Kästner and Miltner, 2018; Miltner et al., 2012; Probandt et al., 2018; Schlüter et al., 2019; Vogel et al., 2014).
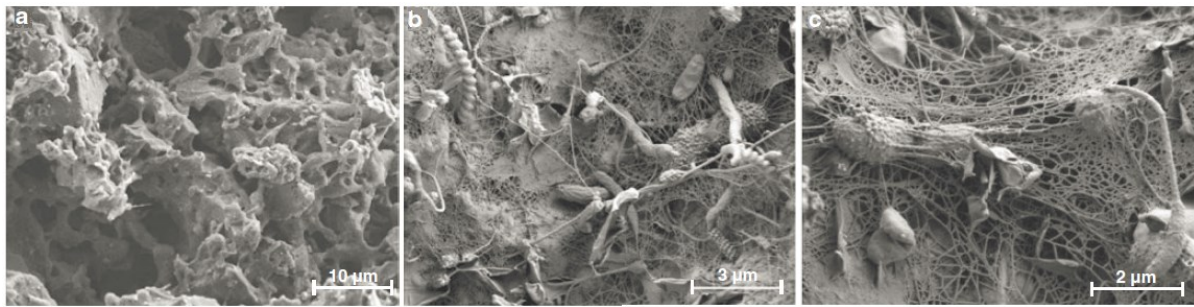
Figure 1-4. Microbial biomass in the soil microverse. Scanning electron micrographs (SEM) of bio-fouled porous beads (diameter ~3- 5 mm) from the seminal data presented by Miltner et al. (2012). For comparison, microaggregates are stable structures with sizes < 250μm and thus two orders of magnitude bigger than bacterial cells shown here. The sterile beads were exposed to natural aquifer water and showed similar biofilm development as observed in soil samples. a) Inner surface of a non-exposed bead; b) colonized bead surface after exposure. Spider-web like structures are extracellular polymeric substance (EPS), which forms web-like structures during sample preparation (critical-point drying); c) colonized bead surface with native cells, empty cell envelopes and their fragments, and much EPS; this is direct evidence of an active microbial loop (only living cells produce EPS; Miltner et al., 2012).
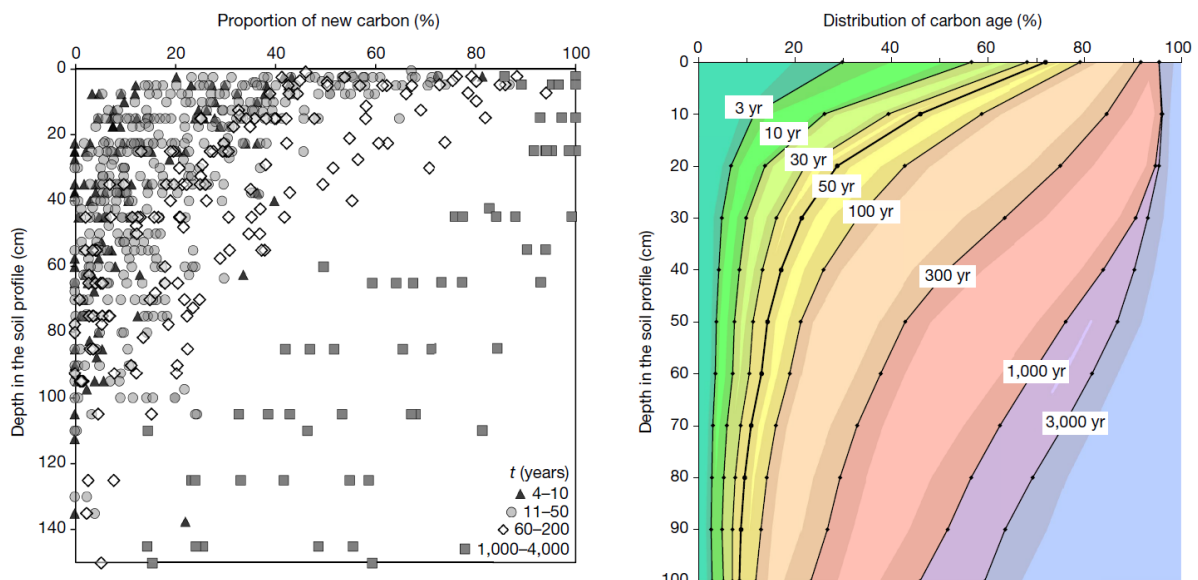
Figure 1-5. Renewal time estimates of soil organic carbon from Balesdent et al. (2018). Left panel: Percentage of new carbon in different soil depths of 112 vegetation change experiments running for time t. Four age classes are shown. Right panel: Meta-Analysis of soil organic carbon age distribution of 55 tropical soils under vegetation change (Description by the authors: proportion of carbon aged less than time t was fitted by a bi-exponential regression of t; grey bands represent ±1 standard error of the estimated mean).

Intriguingly, organic matter associated with such "microbial" microenvironments (small sized particles and pores) shows lower turnover, as compared to organic C associated to larger particles (sand, or particulate organic matter; Gleixner, 2013; Hicks Pries et al., 2017; Marschner et al., 2008; Wynn, 2007). Microsite SOM is also composed of seemingly labile organic compounds, such as carbohydrates, lipids and proteins, supporting its microbial origin (Gleixner, 2013). Relatively strong concentration of microbial cells and their remains in small

pores, and their slow movement thus suggest that microbes may contribute largely to SOM formation, but not necessarily to its distribution and stabilization (Figure 1-2e, f). In contrast, soil invertebrates play a key role in preparing SOM breakdown, distribution and mixing of minerals, mucilage, extracellular polymeric substances (EPS), plant seeds, and microorganisms (Eisenhauer et al., 2019; Filser et al., 2016; Kuzyakov and Blagodatskaya, 2015; Vidal et al., 2019); this way they likely contribute to the long-term stabilization of microbial SOM in micro-aggregates < 250µm (Addiscott, 2010; Six et al., 2004; Totsche et al., 2018). Consequently, SOM renewal slows down with depth: While C younger than 10 years makes up on average 50 % of SOC in the first 10 cm, this value vanishes rapidly to a constant 10 % of SOC beneath 30 cm (Figure 1-5; Balesdent et al., 2018; Castellano et al., 2015; Scheibe et al., 2015). These results show that there is constant but decreasing exchange with young C with depth (Figure 1-3b; Kaiser and Kalbitz, 2012).

Carbon concentrations rapidly decline with depth (Balesdent et al., 2018; Gross and Harrison, 2019) and are thus correlated with slower SOC turnover (Don et al., 2013; Gleixner, 2013; Woolf and Lehmann, 2019). The probability of a decomposer meeting a molecule is thus the easiest explanation for the paradox of apparently old deep-soil SOM (Sierra et al., 2018). Long-term C storage in vast "deserts" of bulk soil, especially at depth, may then be due to preservation through simple "disconnection" (Balesdent et al., 2018; Baveye et al., 2018; Torres-Sallan et al., 2017). However, manifold factors have been shown to influence SOC stocks and DOC properties in the field and under lab conditions. For example, the observation that experimental warming or substrate addition can increase decomposition ("priming") of seemingly "stable" SOM, especially outside of hotspots, questions classical theories of SOM preservation (such as mineral protection) in low-activity regions of soil (Aye et al., 2018; Heitkötter and Marschner, 2018; Hicks Pries et al., 2017; Perveen et al., 2019; Woolf and Lehmann, 2019). Much similar, it is likely anoxia that prevents decomposition within micro-aggregates (Figure 1-2e; Borer et al., 2018; Pitumpe-Arachchige et al., 2018; Six et al., 2004). Slightest changes in boundary conditions such as vegetation input, climate, or soil pH can thus shift the local "equilibrium" and induce changes in DOM composition and transformation from the micro- to the macroscale (Camino-Serrano et al., 2016; Drake et al., 2018; Finstad et al., 2016; James et al., 2019; Kaiser et al., 2001, 2015; Kang et al., 2018; Noacco et al., 2019; Raymond and Saiers, 2010; Roth et al., 2015; Sanderman et al., 2009; Tunaley et al., 2016). In this context, DOM emerges as a rich information source due to its high mobility and fast response to system changes.

1.2.2   Production, transport and transformation of soil DOM

Now, how can we explain soil DOM composition theoretically? The recent cascade model of Kaiser & Kalbitz (Kaiser and Kalbitz, 2012; Leinemann et al., 2018) suggests a differentiation between a "fast-circuit" SOM metabolism, mostly linked to DOM release, and a "slow-circuit" SOM metabolism leading to the formation of SOM associated to smaller sized particles and pores, which builds up over time and would also be stabilized through micro-aggregate occlusion, as explained above (Heitkötter and Marschner, 2018; Kuzyakov and Blagodatskaya, 2015). The cascade model proposes constant exchange reactions between percolating DOM and surface-associated SOM at "hotspot-like" interfaces during downward transport (Kaiser and Kalbitz, 2012; Kuzyakov and Blagodatskaya, 2015), leading to degradation of plant material and phenols, and apparent increases in ^{14}C age, N content, carbohydrates and microbial compounds. However, the mechanisms of this DOM signature shift are unknown.

Decomposed SOM in topsoil "dominant surface layers", rather than recent litter, seems to be key for the sustained DOM production under natural precipitation conditions and fuels belowground foodwebs and recycling (Lee et al., 2018; Lehmann et al., 2018; Leinemann et al., 2018; Malik and Gleixner, 2013; Scheibe and Gleixner, 2014). Consequently, a large fraction of DOM leaching from topsoil is retained in subsoil through mineral interactions such as sorption. Mineral surfaces in natural systems are however rarely pristine and may sorb less DOC than expected (Guggenberger and Kaiser 2003): There is a constant flux of DOM running through soil that likely impregnates available surfaces, promotes biofilm growth and thus, respiration (Kaiser and Kalbitz, 2012; Scheibe et al., 2015). Constant renewal of sorption sites, by, e.g. aggregate turnover, could however promote sustained sorption and be a potential link to the "slow-circuit" part of SOM dynamics (Six et al., 2004; Totsche et al., 2018). With increasing depth, DOC levels become ever lower, and reflect simultaneous decline in SOC stocks and composition (Roth et al., accepted); changes in DOM signatures also reflect the ever slower carbon turnover that likely is both the cause and effect of "disconnected" decomposers and metabolites (Baveye et al., 2018; Sierra et al., 2018).

Cell uptake across membranes requires soluble products with molecular weights < 600 Da (Woolf and Lehmann, 2019). Fungal extracellular ligninolytic enzymes and polysaccharide hydrolases ("exo-enzymes") are thought to play a key role in this step and are found associated to particles rather than in soil solution (Baldrian and Štursová, 2011; Schulze et al., 2005).

Studies indicate that fungal degradation of organic matter leads to simultaneous consumption of colored organic compounds and synthesis of new compounds leading to a mass shift in DOM molecules (Waggoner and Hatcher, 2017; Zavarzina et al., 2018, 2011), and such signatures could become more evident at depth were no "overprinting" by highly soluble decomposed plant material occurs (Kaiser and Kalbitz, 2012; Kästner and Miltner, 2018). Consequently, such shifts in color and molecular weight of DOM have also been observed along soil depth profiles but are not well understood mechanistically (Roth et al., under review). Soil texture, i.e., the particle and pore size distribution of a soil, emerges as a main determinant of SOM/ DOM inputs and transformation. Soil texture affects both carbon preservation (habitat structure; dominance of "inaccessible" microsites) and the flow of water (contact time of soil and water). However, climatic conditions control the amount of water that enters soil, and thus have a top-down impact on C input, recycling and DOM transport (Alvarez-Cobelas et al., 2012; Webb et al., 2018). For example, drought events have been shown to increase soil water repellency, likely due to layer-like assemblage of organic solutes at particles surfaces (Aufdenkampe et al., 2001; Chassé et al., 2015; Kleber et al., 2007; Krueger et al., 2018; Lutfalla et al., 2019; Newcomb et al., 2017; Sollins et al., 2009). In order to disentangle such effects, ecosystem studies linking boundary conditions and molecular DOM properties are needed.

1.3   Chemical characterization of terrestrial dissolved organic matter (DOM)

1.3.1   Isolation of terrestrial DOM

Isolation of DOM from water samples is usually performed by ultrafiltration (UF), reverse osmosis (RO), reverse osmosis combined with electro-dialysis (RO/ED), or solid phase extraction (SPE; Minor et al., 2014; Sandron et al., 2015). Freeze-drying is an option only when samples are low in inorganic salts; then, also direct analysis of water samples may be possible (Sleighter et al., 2009). The SPE method proposed by Dittmar et al. (2008) using PPL resin (Agilent, functionalized styrene–divinylbenzene) is commonly used for the isolation of DOM, and has replaced a former protocol based on XAD-8 and XAD-4 resins (Rohm and Haas, acrylic ester resin and styrene–divinylbenzene resin; Aiken et al., 1992). UF and RO/ED methods are often more timely and require extended equipment and cleaning, and are often reported to be less reproducible than SPE (Green et al., 2014b; Minor et al., 2014; Sandron et al., 2015). Comparisons among SPE sorbents usually show that the PPL resin performs best in terms of DOM recovery and initial sample properties (Dittmar et al., 2008; Green et al., 2014b;

Li et al., 2017, 2016b; Raeke et al., 2016). Li and coworkers improved Dittmar's PPL method by showing that DOC loading affected the DOM composition as assessed by ultrahigh resolution mass spectrometry (FTMS): At higher loading, lower proportions of highly-oxygenated compounds were retained. The authors reached 89% DOC recovery by optimizing loading (1:800, DOC vs. PPL, based on mass) and also showed that, at similar DOC loading, water volume or flow rate did not affect the molecular analysis. Furthermore, RO/ED and original samples were found to be highly similar to the PPL extract in terms of NMR (nuclear magnetic resonance) and FTMS data (Li et al., 2016b). In a second study of the same authors, PPL ranked highest in recovery for a bog (80%) and a marine sample (40%) in comparison to 23 other sorbents used under similar conditions (Li et al., 2017). The authors found that non-polar sorbents (such as PPL and C18) performed best in comparison to other mixed mode, polar or ion exchange sorbents. Raeke et al. (2014) showed that poorly retained standard compounds on PPL did also indicate low response in negative-mode electrospray ionization (ESI) coupled to FTMS detection, suggesting no additional loss in terms of analytical window. In line with their findings, also other studies show that biopolymers (carbohydrates, peptides, lipids) and hydrolysable tannins are not retained well by commonly used non-polar SPE resins (Chen et al., 2016; Liu et al., 2011). Further separation or fractionation of isolated DOM has been largely achieved by gas chromatography (GC) and liquid chromatography (LC), sometimes in 2D setups (Sandron et al., 2015). However, most chromatographic techniques suffer from the high complexity of DOM and produce no sharp peaks due to intrinsic averaging (Brown et al., 2016).

1.3.2 Advanced instrumental analysis of terrestrial DOM

The main techniques used for high-resolution analyses of DOM today are 1) fluorescence spectroscopy, 2) nuclear magnetic resonance spectroscopy, and 3) ultrahigh resolution mass spectrometry (Minor et al., 2014). Optical spectroscopy techniques are usually relatively fast and sensitive but provide limited resolution when applied to complex mixtures. Besides UV, Vis, and IR (ultraviolet, visible light, infrared) spectroscopy, three-dimensional excitation–emission matrix (EEM) fluorescence spectroscopy is the main form of modern optical spectroscopy applied to DOM (Bianchi and Canuel, 2011; Murphy et al., 2018). A drawback of the method is that the chosen parallel factor analysis (PARAFAC) model strongly influences the statistical result of the analysis (Murphy et al., 2018). Fluorescence spectroscopy is also highly selective for fluorescing conjugated π-electron systems that only represent ~ 1% of the total DOC ("FDOM"; Leenheer, 2009; Murphy et al., 2018). In contrast, in nuclear resonance

spectroscopy (NMR), all magnetic nuclei of an atom such as ^{1}H and ^{13}C absorb energy at a specific pulsed radiofrequency related to nucleus-specific energy level transitions, and subsequently interact with a strong external magnetic field by relaxation (Keeler, 2011). The specific absorbed frequency yields quantitative and non-destructive structural information on nuclei of interest (Hertkorn, 2014; Minor et al., 2014). NMR shows a high degree of signal overlap when applied to complex mixtures such as liquid-state DOM and solid-state SOM (Hertkorn, 2014; Minor et al., 2014). The high resolution that can be achieved is also outweighed by the method's insensitivity, i.e., it requires large amounts of material, and involves long measurements (less so for ^{1}H NMR). Specially tailored experiments (2D, 3D) are needed to "extract" information on specific substructures of interest (Bell et al., 2015; Hertkorn, 2014). This way, substantial contributions regarding substance groups in DOM have been made in recent years, providing evidence for the presence of previously unknown fractions of carboxyl-rich alicyclic molecules ("CRAM", Hertkorn et al., 2006), material derived from linear terpenoids ("MDLT", Lam et al., 2007), oxidized sterols (Woods et al., 2012) and carotenoids (Arakawa et al., 2017).

1.3.3 Ultrahigh resolution mass spectrometry

Ultrahigh resolution mass spectrometric (FTMS) techniques allow for the unsurpassed separation (resolution) of complex mixtures by mass and charge and have emerged as indispensable tools to study molecular detail in marine and aquatic DOM (Hertkorn et al., 2013, 2008). In the field of soil science, they are not yet widely applied. The FTMS measurement is usually fast (~ 10 min per sample), and does not require large amounts of sample, in strong contrast to NMR spectroscopy (Hawkes et al., 2016; Hertkorn et al., 2013). The only prerequisite for ultrahigh-resolution mass separation is ionization, which is required for separation by mass and charge (*m/z*) in static or dynamic electric or magnetic fields (Gross, 2011). "Soft" atmospheric pressure ionization (API) techniques such as electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) have opened the field of mass spectrometry-based "omics" technologies as they are ionizing molecules in their non-fragmented state and allow analysis of compounds previously unamenable to MS, e.g. proteins, or complex mixtures, such as DOM (Gross, 2011). Negative mode-ionization is usually preferred for DOM analysis due to the dominance of oxygen functionalities (Hertkorn et al., 2008). Modern FTMS instruments such as Fourier ion cyclotron resonance mass spectrometers (FT-ICR MS) or Orbitrap mass spectrometers allow the simultaneous detection of thousands

of peaks that can be assigned a molecular formula due to very high mass accuracies < 1 ppm (Table 1-1; Marshall and Hendrickson, 2008; Perry et al., 2008; Zubarev and Makarov, 2013). The formula assignment is based on combinatorial calculations using the known exact masses of the different isotopes expected in DOM (e.g., ^{1}H, ^{12}C, ^{13}C, ^{14}N, ^{15}N, ^{16}O, ^{31}P, ^{32}S, ^{34}S, etc.), and requires information on the charge state of the ion; depending on the mass error of the instrument, unlikely combinations of atoms can be ruled out (Kind and Fiehn, 2007). The FTMS techniques show the highest potential to isolate and putatively identify individual constituents in complex mixtures such as DOM because of their versatility: In general, MS can be coupled to various forms of liquid and gas chromatography (LC, GC) and ionization techniques (ESI, MALDI, chemical ionization, etc.), and allows further experiments upon ionization (i.e., fragmentation, gas-phase reactions, ion mobility), or even preparative steps through custom-made soft-ion landing devices (Arakawa et al., 2017; Baumeister et al., 2018; Brown et al., 2016; D'Andrilli et al., 2010; Gaspar et al., 2009; Gologan et al., 2004; Johnson et al., 2016). Especially the possibility to isolate and fragment intact precursor ions offers manifold avenues for the analysis of structural detail (i.e., indicative product ions, neutral losses or rearrangements; recurring mass difference patterns; comparison with databases; Brown et al., 2016; Petras et al., 2017).

The Orbitrap analyzer has emerged as an alternative to the high-performance FT-ICR MS instruments (Table 1-1; Hawkes et al., 2016). Orbitrap instruments achieve resolving powers greater 100.000 and have been used extensively to study DOM, petroleum, and other complex natural mixtures (Galindo and Del Nero, 2015; Hawkes et al., 2016; Mangal et al., 2016; Zhurov et al., 2013). Although resolution of specific elemental combinations (especially involving N, S and P) is often precluded, major oxygen-containing (CHO) and oxygen- and nitrogen-containing (CHNO) formulae can be detected easily. Novel instruments allow for extended transient lengths, and achieve resolving powers greater 500.000 < 300 *m/z*, thus widening the analytical window to also detect NSP formulae. These improvements and the much higher availability/ lower cost of Orbitrap instruments explain the fueled interest in their application for DOM studies worldwide (Hawkes et al., in preparation). However, despite an overall similarity of data obtained by FT-ICR MS and Orbitrap instruments (e.g., Hawkes et al., 2016), there is still much unknown about signal variability – especially spectrum shape – among instruments (Hawkes et al., in preparation).

Table 1-1. Comparative features of both analyzers (compiled from Marshall and Hendrickson, 2008; Perry et al., 2008; Zubarev and Makarov, 2013).

	FT-ICR MS instruments	Orbitrap instruments
Image (from manufacturer)		
Size of the instrument	High due to superconducting magnet	Theoretically small, benchtop and portable instruments possible
Flexibility of the analyzer	High due to storage of trapped ions in the ICR cell	Low, no storage of ions; depending mainly on front end
Costs	Very high instrumental costs High maintenance costs	High instrumental costs Low maintenance costs
Instruments used in this thesis	SolariX XR (introduced 2009; mod. 2013; Bruker Daltonik GmbH, Bremen, Germany)	Orbitrap Elite (introduced 2011; Thermo Fisher Scientific, Bremen, Germany & Waltham, USA)
Resolution	~ 1.000.000, up to 2.000.000 possible with modified XR cell This thesis: ~ 615.000 @ 400 *m/z*	~ 150.000; Up to 480.000 possible with beta software; Chapter 3: ~ 220.000 @ 400 *m/z* Chapters 4 and 5: 480.000 @ 400 *m/z*
Accuracy in mass determination	< 1 ppm, < 0.1 ppm possible with modified XR cell sub-ppm also by internal calibration	1 – 5 ppm < 2 ppm with internal calibration
Stability of mass accuracy	Stable for months	Stable for days to weeks
Field applied for trapping	Magnetic	Electrostatic
Ion motion used for detection	Cyclotron rotational frequency	Axial oscillation frequency
Frequency link to m/z	$f \propto \dfrac{1}{m/z}$	$f \propto \dfrac{1}{\sqrt{m/z}}$

The exact mass measurements are conducted in special detection cells (ICR or Orbitrap cells). In ICR instruments, the ions rotate in a set of opposing electrode pairs within a static magnetic field, whereas in Orbitrap instruments (Figure 1-6), they both rotate and oscillate axially, i.e., "orbit", along an inner spindle-like electrode in a barrel-like outer electrode and within an electrostatic field (Marshall and Hendrickson, 2008; Perry et al., 2008; Zubarev and Makarov, 2013). In these fields, ions move at different speeds depending on their mass to charge (*m/z*) ratio. The frequency spectrum obtained from ions of similar *m/z* orbiting in phase as coherent "ion packages" can then be back-calculated by Fourier transform to yield the initial *m/z* and ion abundance of all ions (molecules), i.e., the conventional mass spectrum. It is important that the ions move coherently to allow detection based on frequency: Only then, ion packages will

induce currents in opposite poles of the detection cells that relate to their rotation frequency (and thus, *m/z*) and abundance. In FT-ICR MS, the coherence of ion packages is induced by a frequency pulse (excitation) to make full use of the ion cyclotron frequency; in the Orbitrap, ion injection is controlled by a specific C-trap designed for the fast injection of a tight ion package that leads to coherent axial oscillation of different ion clouds upon measurement in the detection cell without an additional excitation step (Perry et al., 2008). Isolation of precursor ions prior to exact mass determination can be conducted in the front-end ion trap (Velos Pro MS in Figure 1-6) of the Orbitrap.
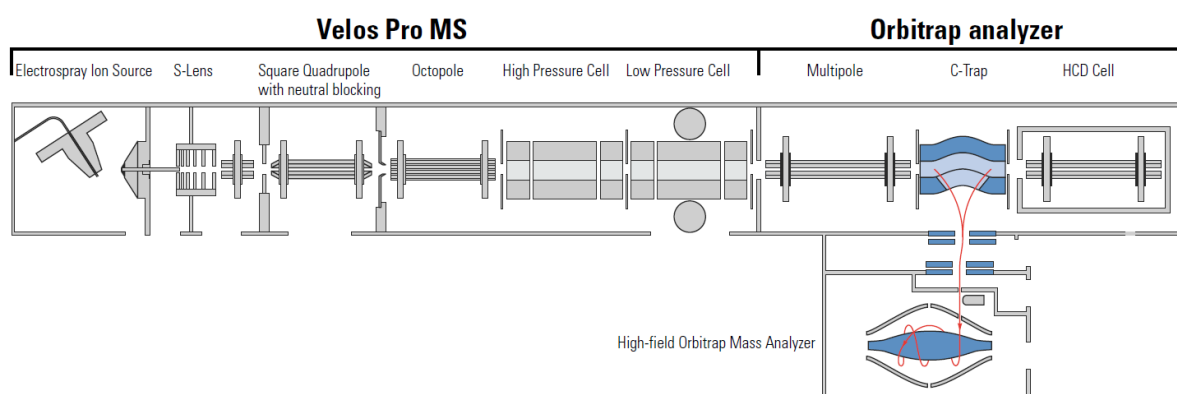
Figure 1-6. Schematic view of the Orbitrap Elite (Thermo Fisher Scientific, 2013). ESI source, left, used for ionization of liquid sample and source-induced dissociation (SID); ion trap, "high pressure/ low pressure cell"; used for collision induced dissociation (CID) by Nitrogen gas and low-resolution mass measurements; Higher charge collision dissociation (HCD) with no low mass cutoff, right. Lower right, beneath "C-trap" used for controlled injection of ions into Orbitrap analyzer; Orbitrap analyzer used for ultrahigh resolution mass measurements. Flight path of ion shown in red, orbiting motion around the central spindle-like electrode; surrounded by "barrel-like" electrode used for frequency detection (split in half along rotation axis). S-Lens, ion funnel; remaining quadrupoles and octopoles: part of ion optics for ion beam focusing.

Despite the high potential of FTMS techniques, there are also some important drawbacks that have to be noted. Ion abundance is not necessarily equal to concentration; it may be influenced by the affinity of a molecule to be ionized in the first place, and ion suppression in detection cells can also modify the signal (Marshall et al., 1998; Rodgers et al., 2019). The length of the measurement (transient length, in the range of hundreds of ms) determines the achieved mass resolution, but may be limited by chromatographic peak widths. FT-ICR instruments generally achieve higher resolutions (Marshall and Hendrickson, 2008). The ultrahigh resolution of FTMS instruments and their ease of use (flexibility, simple injection, optional automation and hyphenation, linear range, fast and reproducible measurement) are outweighed by some drawbacks that have to be addressed for each application. In non-targeted DOM applications, these drawbacks mainly arise from the complexity of the material. The electrospray ionization

process is highly selective and still not fully understood (Nebbioso et al., 2010; Raeke et al., 2016), even more so when thousands of molecules, and many more structural isomers, compete for charge at the same time (Hawkes et al., 2019; Rodgers et al., 2019). This also complicates the consideration of blank peaks and contaminants, as those might by largely suppressed when they compete for charge with an overwhelming number of analytes (Sleighter and Hatcher, 2011). None of the available soft ionization techniques provides universal ionization, and many studies have shown that their complementary use increases the analytical window (D'Andrilli et al., 2010; Hertkorn et al., 2008; Hockaday et al., 2009; Minor et al., 2014). DOC is usually injected at the same concentration, but this does not exclude yet unknown effects, e.g., of DOC origin (Nielsen et al., 2018): For example, the amount of C that is ionized remains unknown (Fatayer et al., 2018; Rodgers et al., 2019). On top, there exists no "standard" mixture mimicking the complexity and composition of DOM, thus no true quantification is achieved, and comparisons remain semi-quantitative at best. The knowledge of empirical formulae and molecular weights of molecules alone does not allow structural assignments, besides theoretical considerations (aromaticity, heteroatom content). Valuable additional information can be drawn from indicative derivatization or degradation treatments conducted prior to measurement, and from fragmentation experiments in the mass spectrometer (Arakawa et al., 2017; Minor et al., 2014; Pohlabeln and Dittmar, 2015). Finally, the absence of general guidelines for tuning and quality measures – lastly connected to the question of what should be regarded as the "true" mass spectrum of a DOM sample – makes comparisons among different labs tedious, and hampers large scale comparisons (Swenson et al., 2014). Thus, novel approaches to improve data comparability are needed to gain better understanding of the DOM signal seen by FTMS instruments.

1.4   Identification of ecosystem imprints in terrestrial DOM

1.4.1   Identified markers in DOM: State of knowledge

Structures that are identifiable by means of classical lab procedures – isolation, purification, structure elucidation – make up less than ~10 % of the DOC (Minor et al., 2014; Swenson et al., 2015; Thurman, 1985) and encompass, for example, hydrolysis products of polysaccharides and proteins, fatty acids, oxidation products of lignin, and reduction products of terpenoids and carotenoids (Arakawa et al., 2017; Arakawa and Aluwihare, 2015; Minor et al., 2014; Repeta, 2015; Thurman, 1985). The structural elucidation of the large remainder of extracellular DOM has progressed only slowly (Bianchi and Canuel, 2011; Minor et al., 2014): FTMS methods

now reveal the molecular formulae of many unknown metabolites, but these assignments are ambiguous in terms of structure, and thus theoretical assignments based on elemental composition remain preliminary (Davies et al., 2015; Reemtsma, 2010). Several studies have, however, suggested putative structures in DOM fractions, mainly from aquatic and marine settings (Figure 1-7). Organic matter degradation seems to lead to converging, universal, but also indistinguishable forms of DOM across ecosystems (Figure 1-8c, e; Kellerman et al., 2014; Lechtenfeld et al., 2014; Roth et al., 2014). FTMS fragmentation data suggest that the underlying DOM structures may be caused by multiple isomers at dilute concentrations (intrinsic averaging; Zark et al., 2017; Zark and Dittmar, 2018). This finding has been supported by prior separations through chromatography (Brown et al., 2016; Hawkes et al., 2018a), isotope exchange (Kostyukevich et al., 2014; Stenson et al., 2014) and ion mobility (Gaspar et al., 2009; Leyva et al., 2019; Lu et al., 2018). Despite this vast chemodiversity and "universality" of DOM revealed by FTMS methods, large differences between DOM from various ecosystems are frequently reported within the literature (Figure 1-8; Rossel et al., 2013). These differences are often related to small and "rare" signals, while signals detected ubiquitously tend to be of high abundance (Kellerman et al., 2014; Roth et al., 2014). Significant source or process imprints of such unique or ubiquitous signals may serve as valuable new markers (Figure 1-8a, b). For example, experimental treatments such as bacterial or photochemical degradation and sorption show clear imprints in DOM.
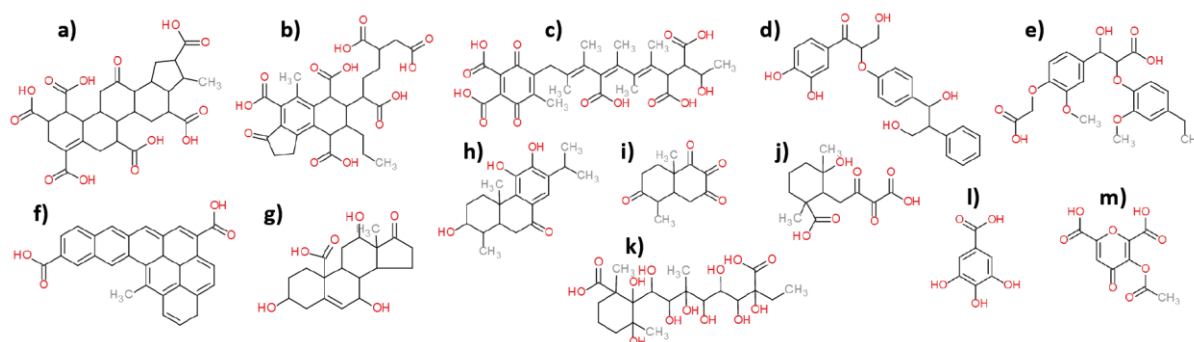
Figure 1-7. Suggested structures of molecules present in DOM. Alicyclic structures dominate; DOM of land plants is linked to lignin decomposition. a – c) CRAM structures from marine DOM (Hertkorn et al., 2006); d) Lignin oxidation product from SRFA (Stenson et al., 2003); e) modified lignin structure from a water extract of wood decomposed by brown-rot fungi (Liu et al., 2011); f) modified polyaromatic hydrocarbon structure from the abyssal ocean (Dittmar and Koch, 2006); g) oxidized sterol from SRNOM (Woods et al., 2012); h – j) alicyclic terpenoid structures from SRFA (Arakawa and Aluwihare, 2015); k) part of a oxidized carotenoid structure from marine DOM (Arakawa et al., 2017); structures h – k in part also suggested elsewhere for lake DOM (Lam et al., 2007); l) gallic acid exemplarily shown as a base unit of hydrolysable tannins; m) proposed structure from Dictionary of Natural Products (Chassagne et al., 2019), acetylated 3-Hydroxy-4-oxo-4H-pyran-2,6-dicarboxylic acid, for the molecular formula $C_9H_6O_8$ which is a potential forest marker in soil DOM (Roth et al., 2014). Formation of alicyclic structures from lignin oxidation products is discussed by Waggoner et al. (2015).
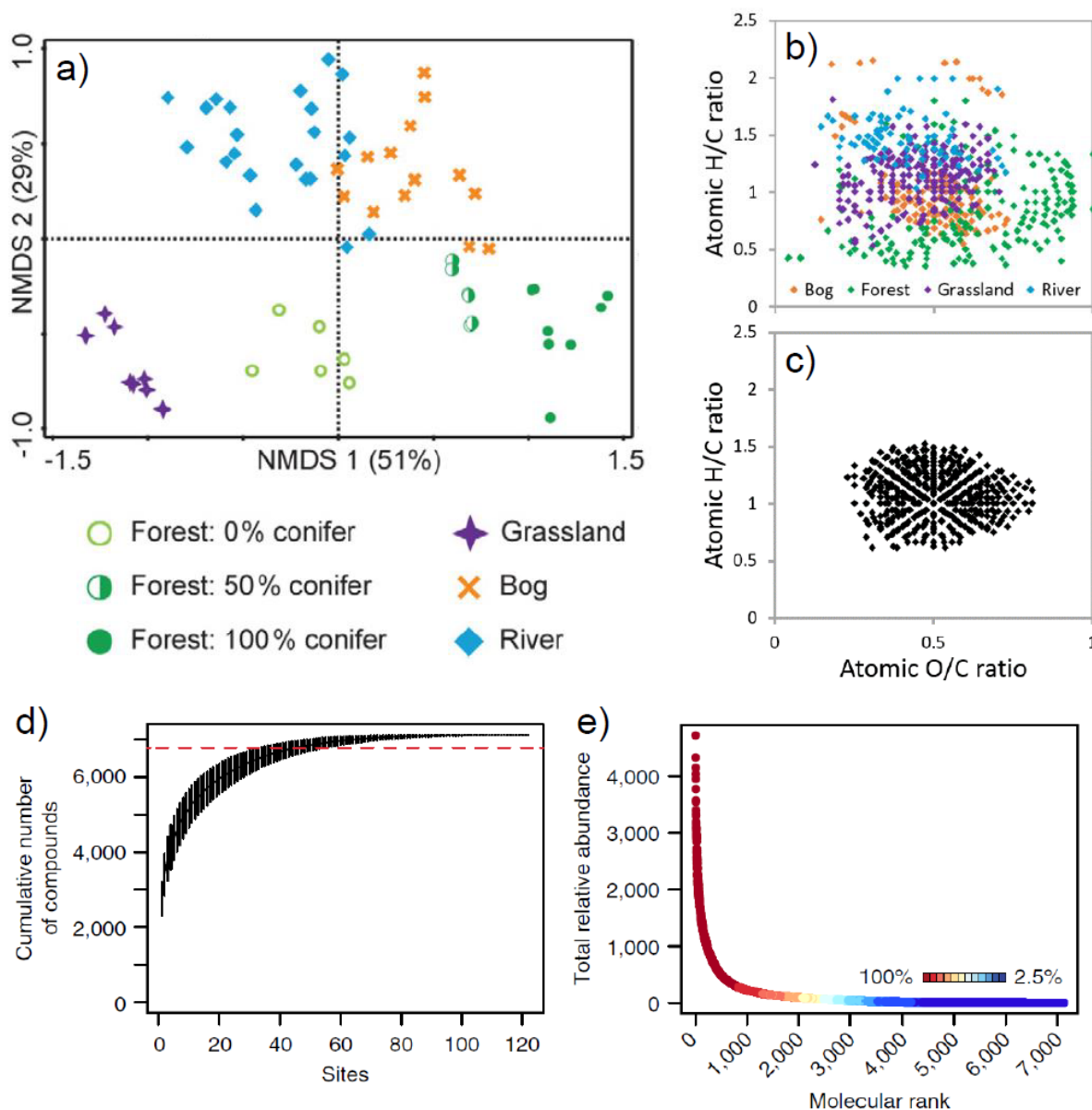
Figure 1-8. Ecosystem-specific markers in DOM identified by FTMS. Modified from Roth et al. (2014) (a – c) and Kellerman et al. (2014) (d, e). a) Non-metric dimensional scaling analysis of 66 DOM samples from different environments, showing clear separation of soil water and surface waters on NMDS 2 and separation according to pH on NMDS 1 (conifer forests = lowest pH). b) Unique formulae of different sample sets from a, shown in a Van Krevelen plot that shows molecular formulae according to their atomic ratios of H/C and O/C. c) Set of molecular formulae that were common to all environments. d) Sampling plot of molecular "species" (here, formulae) for a set of 120 lakes distributed across Sweden. With each lake ("site" on the x-axis), novel species are added. Error estimate is confidence intervals (1000 permutations), red line is 95% of molecular formulae. The plot shows that the chemodiversity of lakes in this climate (as seen by FTMS) is reached after sampling ~ 50 lakes and saturates after ~ 80 lakes, showing that the chemodiversity of Swedish lake DOM can be estimated with ~ 7000 unique molecular formulae. e) Rank-abundance plot of species (formulae) shows that ubiquitous species (color code, occurrence in % of samples) are also most abundant (in terms of relative abundance, averaged). This finding agrees with panel c and other studies (Flerus et al., 2012): this region of the Van Krevelen plot is connected to peaks with highest ion abundances in mass spectra of DOM, but is also most diverse in terms of possible structural isomers (Hertkorn et al., 2007).

It however becomes clear that these signals are far from being constrained:

1) Differentiation is limited to broad annotations ("aliphatic", "aromatic", "nitrogen-containing") that show overlap; e.g., aromatic formulae respond to microbial and photochemical degradation, and sorption (Galindo and Del Nero, 2014; Kujawinski et al., 2004; Riedel et al., 2013; Rossel et al., 2013; Stubbins et al., 2010; Ward et al., 2013).

2) Experiments show deviating results depending on sample origin (and lab?) and rarely account for interactive effects. For example, photo-produced formulae differ among staring materials (Rossel et al., 2013; Stubbins et al., 2010), and the combination of two factors results in a synergistic molecular picture (Chen et al., 2014; Rossel et al., 2013).

3) Molecular dynamics differ among ecosystems and scales. For example, molecular formulae related to preserved DOM differ between environments (Drake et al., 2018; Kellerman et al., 2015; Lechtenfeld et al., 2014); N-containing compounds are abundant at higher temperature in boreal lakes whereas in a boreal river and temperate soils they are abundant at colder temperatures (Kellerman et al., 2014; Roth et al., 2015, 2013).

4) The responding molecular formulae are rarely further identified by complementary methods (Boiteau et al., 2019; Kügler et al., 2019; Minor et al., 2014; Roth et al., 2014).

A different body of studies analyzes whole-ecosystem-specific DOM composition along with environmental factors that likely control differentiation, beyond single processes (Dubinenkov et al., 2015; Gonsior et al., 2016; Hawkes et al., 2018b; Hertkorn et al., 2016; Jaffé et al., 2012; Kellerman et al., 2014; Kothawala et al., 2015; Kujawinski et al., 2009; Mosher et al., 2015; Ohno et al., 2010; Riedel et al., 2016; Roth et al., 2014). Although offering novel hypotheses on causal explanations of the observed changes, none of these studies has followed up by elucidating actual marker structures.

At this point, the research community thus is making insufficient use of the actual analytical potential of the FTMS analyzer in both applications aiming to identify and/ or compare signals (Davies et al., 2015; Petras et al., 2017; Reemtsma, 2010; Swenson et al., 2014). For example, Roth et al. (2014) reported on a set of forest markers and proposed a tannic origin, as based on their elemental composition (highly oxygenated, O/C > 0.6; relatively low saturation, H/C < 1; Figure 1-8b). Bog and forest soil DOM has been repeatedly attributed to aromatic and phenolic constituents; this, however, seems to be mainly a pH effect (Adamczyk et al., 2018; Gallet and Pellissier, 1997; Gonsior et al., 2016; Roth et al., 2015, 2014, 2013; Simon et al., 2019; Soucémarianadin et al., 2017; Strobel, 2001; Zwetsloot et al., 2018). Their tannic origin has to

be tested by complementary methods such as fragmentation (tandem MS). Previous attempts to elucidate DOM structures by tandem MS have, however, resulted in the notion that mainly non-informative fragments such as CO_2 and H_2O were found (Brown et al., 2016; Hawkes et al., 2018a; Witt et al., 2009; Zark and Dittmar, 2018), and also structure databases show generally less than 5% hits due to the high number of unknowns present in DOM (Petras et al., 2017; Zhang et al., 2014). Novel approaches to study source-indicative imprints are thus needed to better exploit tandem MS data of complex DOM samples and open new avenues in marker identification.

1.4.2 The Rio Negro: A natural laboratory to study ecosystem markers
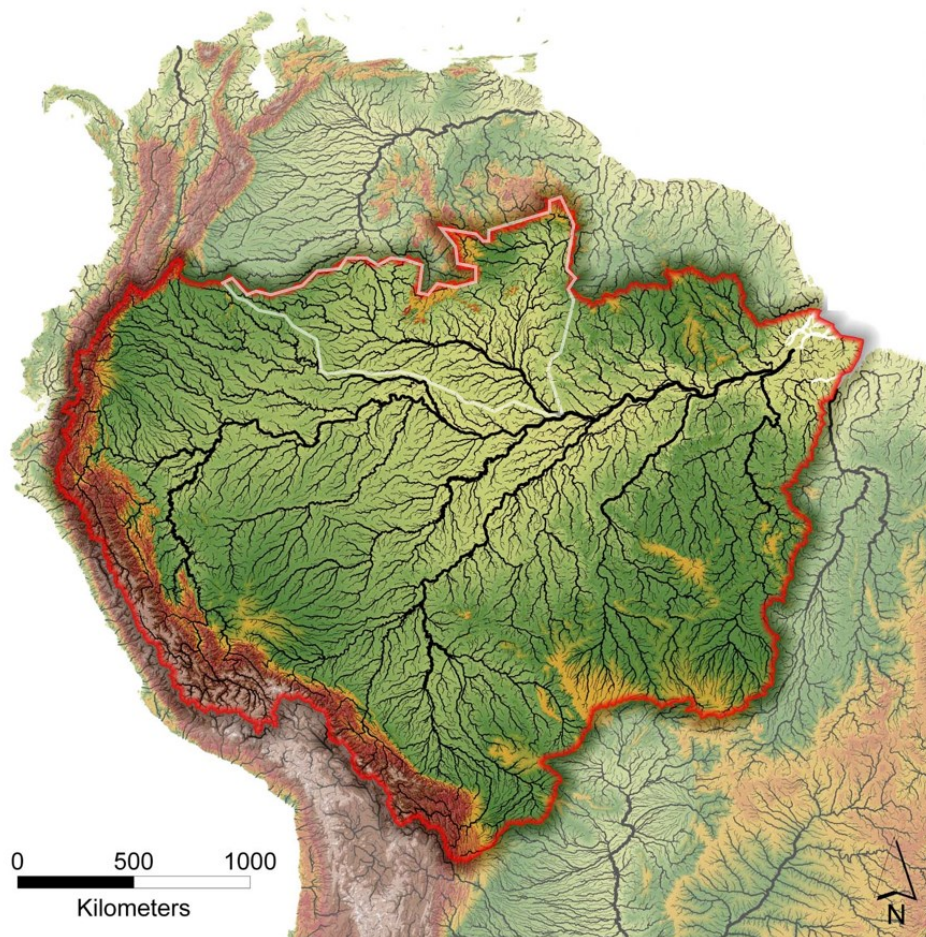
Figure 1-9. The Amazon basin. Data from the HydroSHEDS project, USGS (https://hydrosheds.cr.usgs.gov). Only major water bodies are visualized; line width is proportional to basin area upstream. The Rio Negro basin is marked with a white line.

The Rio Negro watershed is used herein as a model system to apply developed methods for the study of DOM molecular composition. Tropical forests are an important piece of the global climate system and its interconnected elemental cycles (Nagy et al., 2016; Townsend et al., 2011). The Amazonian rainforest (Figure 1-9) stores vast amounts of fixed C in aboveground biomass (~110 Gt-C or 25% of global plant biomass; Bar-On et al., 2018; Gloor, 2016) and soils (~70 Gt-C in SOC up to 1 m depth; Batjes and Dijkshoorn, 1999; Cerri et al., 2007) that are prone to release due to increasing deforestation and frequency of fires in the course of global change (Aragão et al., 2018; Cerri et al., 2007; Gloor, 2016; James et al., 2019; Silvério et al., 2019). Amazon soils are deeply weathered and thus export only low amounts of cations, in contrast to Andean soils and rivers (Horbe et al., 2016; Markewitz et al., 2001).

The Rio Negro is the world's largest blackwater tributary in terms of mean annual discharge and the largest tributary of the Amazon. It covers 12% (700.000 km²) of the basin's area (Frappart et al., 2008). The river is known for its low pH (~ 4) and high export of DOC related to the high abundance of natural phenolic acids (Ertel et al., 1986; Furch and Junk, 1997; Gonsior et al., 2016; Simon et al., 2019), pointing to the tight cycling of nutrients (N, P; Townsend et al., 2011). Despite DOC, large amounts of exported silica, iron, and aluminum indicate active podsolization processes within the watershed (Figure 1-10; Do Nascimento et al., 2008; Guinoiseau et al., 2016; Patel-Sorrentino et al., 2007). The podsolization process is promoted initially by high precipitation and poor drainage (Bravard and Righi, 1990). Gradual clay impoverishment through penetration of highly acidic topsoil water rich in DOM promotes the development of reducing conditions upon DOC oxidation (Do Nascimento et al., 2004; Lucas et al., 2012). Such conditions favor the further dissolution of minerals and are amplified by the high complexing capacity of DOM (Ishida et al., 2014) as well as good drainage properties of aggregates of Fe- and Al oxides (Demenois et al., 2018; Lehmann et al., 2001). The high saturated hydraulic conductivity of leached soil layers allows vertical and lateral expansion and further leaching of Fe oxides and kaolinite while quartz remains, thus leading to sandy deposits in topsoil (Do Nascimento et al., 2008; Ishida et al., 2014; Lucas et al., 2012; Patel-Sorrentino et al., 2007). Fe and Al are preferentially leached as organic complexes, and precipitated/ remobilized in deeper soil depending on redox conditions; again, high DOC fluxes promote this mobilization (Do Nascimento et al., 2008; Lucas et al., 2012).
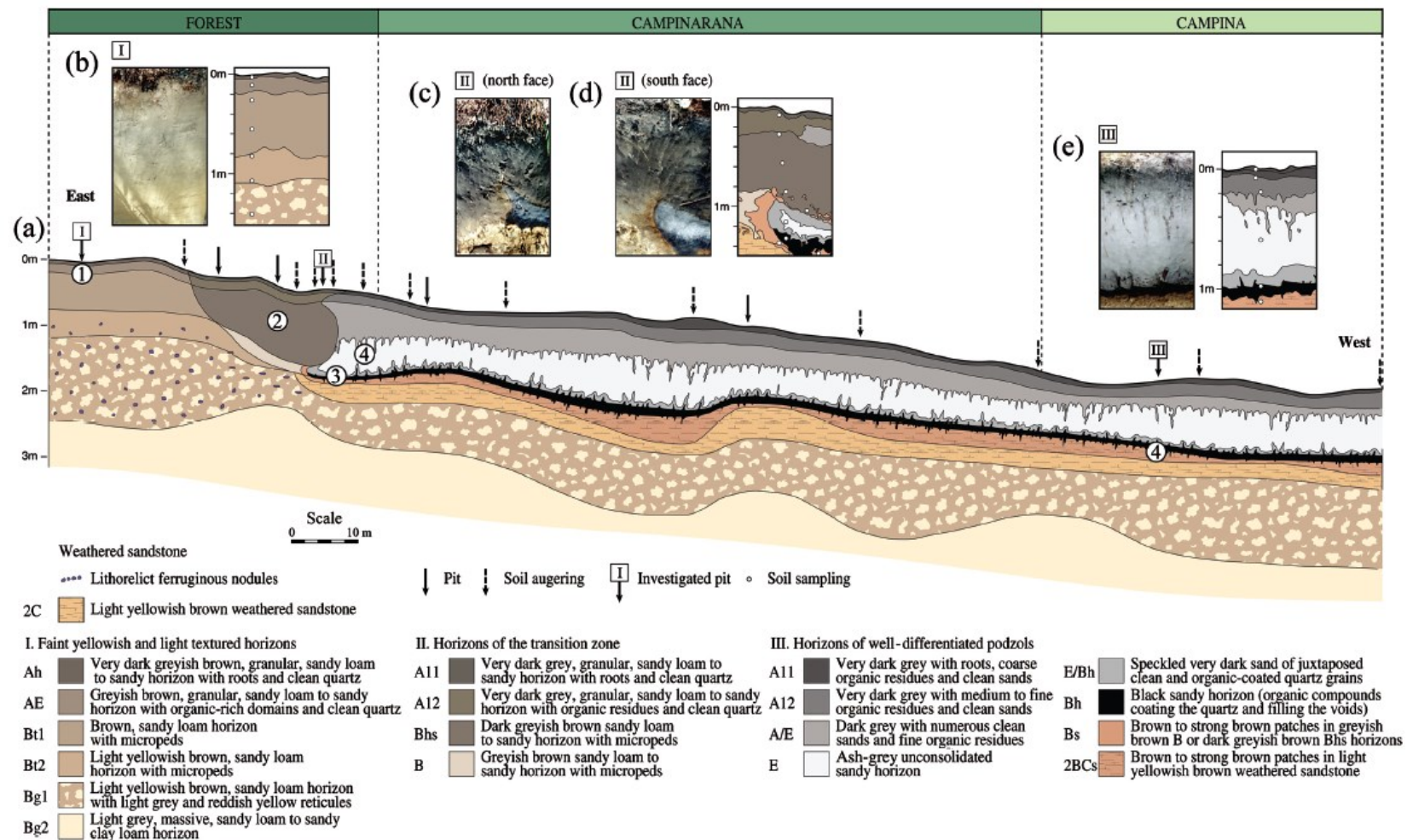
Figure 1-10. Exemplary landscape section of a typical Oxisol - Podzol catena with soil and vegetation description (a). The specific Podzols in this area are hydromorphic, i.e., waterlogged, probably through hardpan formation in the deeper soil (Bh, Bs horizons), and the ongoing upslope transition from Oxisols to Podzols is nicely visible. Photographs of profile and demarcation of horizons in: (b) pit I, (c) pit II (north face), (d) pit II (south face) of the transition zone, (e) pit III within the depression (Do Nascimento et al. 2004). Numbers 1 -4 (in circles) refer to stages of podsolization.

Many authors have drawn a link between podzol occurence (highly acidic whitesand areas) and the blackwater origin of the Rio Negro (Ertel et al., 1986; Goulding et al., 1988; Leenheer, 1980; Sioli, 1954). Whitesand areas do however show clear differences in landscape position (dry plateau/ wet valley) that may have effects on DOM properties (water fluxes; litter turnover; respiration rates; Zanchi et al., 2015, 2014, 2011); also, fine-textured soils (Oxisols) would be expected to show differences in DOM composition as compared to coarse-textured sandy soils (Podzols; Remington et al., 2007). It has been further hypothesized that the molecular imprint of whitesand ecosystems would be traceable through dedicated markers in DOM (Bardy et al., 2011; McClain et al., 1997). This is favored by the dark water color that may cause 1) stratification of large water bodies, favoring oxygen depletion beneath the warmer surface, and 2) low light penetration, thus minimizing algal growth and photochemical decay of DOM (Goulding et al., 1988; McClain and Naiman, 2008; Sioli, 1954). Recently, two studies have reported on watershed-specific molecular markers of the Rio Negro (Gonsior et al., 2016; Simon et al., 2019), showing clear parallels to pH- and DOC-related indicative DOM markers observed also in temperate conifer forest and boreal bog ecosystems (Roth et al., 2015, 2014). In line with previous reports of phenolic DOM constituents, their molecular formulae were characterized by relatively high-molecular weight, low H/C (hydrogen-depleted) and high O/C (oxygen-rich), pointing towards larger polyphenols with multiple aromatic rings. Formaly, these markers could be classified as "tannic" molecules, following similar suggestions based on H/C and O/C ratios (Minor et al., 2014; Roth et al., 2014). Molecular characterization of DOM from whitesand areas and clear identification of phenolic markers is however still lacking in the literature and may open up new possibilities to link watershed-specific markers at different scales.

1.5   Hypotheses and objectives of this thesis

The overall aim of this thesis was to develop novel techniques to identify ecosystem markers in terrestrial DOM by means of Orbitrap mass spectrometry. The research was guided by the following hypotheses:

Hypothesis 1: The Orbitrap is suited for the molecular-level analysis of DOM along its passage through terrestrial ecosystems.

Hypothesis 2: Headwater DOM of Amazonian whitesand ecosystems differing in eco-hydrology (dry/ wet system) show differentiation on the molecular level and link to previously described Rio Negro-specific ecosystem markers (tannin-like, phenolic compounds).

Hypothesis 3: Direct injection tandem MS analysis of soil water DOM reveals the ecosystem imprint of plant-related tannin-like polyphenol structures found in low-pH environments.

1.6   Thesis organization

To evaluate the performance of the most recent Orbitrap analyzer as a more accessible alternative for the molecular-level analysis of DOM in **chapter 3**, we compare the instrument with an established 15 Tesla FT-ICR MS on a diverse suite of 17 DOM samples regarding 1) ion abundance patterns, 2) differential effects of DOM type on information loss, and 3) derived biogeochemical information. The sample set represents a hypothetical soil/ aquatic gradient through a large variety of terrestrial ecosystems; we evaluate the suitability of our approach to improve the reproducibility of DOM analyses among labs.

In **chapter 4**, we apply our Orbitrap method to track ecosystem-specific DOM markers from podzol headwater areas of blackwater streams in the Rio Negro watershed. We compare two contrasting groundwater transects and discuss the imprints of microclimatic conditions in terms of water isotopy, radiocarbon age of DOC, and molecular DOM fingerprints assessed by the Orbitrap. We compare the headwater transect data with published datasets on Rio Negro-specific markers to evaluate possible imprints of headwater markers onto the river's DOM signal.

To reveal ecosystem-specific imprints in terrestrial DOM that go beyond compound group assignments via Van Krevelen plots, we present a novel approach to decipher complex chimeric mass spectra of soil DOM obtained by direct-injection electrospray ionization Orbitrap tandem mass spectrometry (DI-ESI-Orbitrap MS/MS) in **chapter 5**. Pairwise exact mass differences (MDs) of all precursor and product ions were used to derive a tandem MS-based MD matrix that was subsequently matched against two lists of known MDs: 1) literature-known but non-indicative MDs, and 2) highly indicative MDs of a set of 14 phenolic substances obtained from tandem MS experiments on the same instrument. We fragmented four soil DOM isobars (m/z 241 301, 361 and 417) at three CID stages (normalized collision energies of 15, 20, 25) to test differences in DOM fragmentation behavior.

Finally, in **chapter 6**, we synthesize the findings from this work in order to derive overarching conclusions and novel hypotheses for future research focusing at the molecular identification of DOM ecosystem-markers in terrestrial environments, with a focus on the dynamics of the soil environment and signal propagation in fluvial networks and implications for ecosystem marker identification.

All methods applied and materials used are listed in **chapter 2**; supplemental data is provided in the appendix (**chapter 7**). Tables and figures are numbered according to their chapter ("Table 1-1"). Supplemental data is sorted into *tables*, *figures* and additional *notes*. These objects are ordered according to when they are mentioned in the main text ("Table 7-1").

2 Material and methods

2.1 Resolution of molecular detail in DOM by Orbitrap mass spectrometry[1]

2.1.1 Sample preparation

We used a diverse set of samples (Table 2-1), including two frequently used reference samples, a Suwannee river sample provided by IHSS as a powder (International Humic Substances Society, 2R101N, Green et al., 2014a), and a deep sea sample (NEHLA) from North Equatorial Pacific Intermediate Water collected from the Hawaiian deep ocean water well (Green et al., 2014b). The IHSS sample was reconstituted in ultrapure water before extraction. The other samples covered a large gradient of environmental settings and included water samples from bogs, soils, aquifers, lakes, and rivers (Table 2-1). Bog samples N3B and N8B and Yensisej river sample N8R came from a transect study along the Yenisey river (Roth et al., 2013). Forest soil water samples were originally collected over two weeks in November 2005 from permanently installed glass ceramic suction plates (1–1.6 mm pore size) in 5 cm depth from forest sites HS2-5, W1-5 and T-5. Site H2S-5 is a beech forest on a Cambisol located in Hainich National Park, Germany (Tefs and Gleixner, 2012), site W1-5 is located in a spruce forest over a Podzol at Wetzstein, Germany (Kindler et al., 2010), and site T-5 is located in Thann, Germany in a stand of maple trees in a mixed pine and spruce forest (Sachse et al., 2009). These samples were stored as freeze-dried extracts and reconstituted in ultrapure water prior to extraction. Grassland soil water samples were taken from a semi-natural grassland on an eutric Fluvisol, at the Jena Experiment, located in Jena, Germany (Lange et al., 2015). The samples were collected with similar suction plates installed in 10, 20, 30 and 60 cm depth as described above in early May 2014 (Roth et al., accepted). Aquifer samples were sampled from deep wells installed in karstic aquifers in Hainich National Park (Küsel et al., 2016), from an oxic (H3-2b) and anoxic aquifer (H5-3a), in July 2014. Tap water was sampled at the Institute for Geosciences of the Friedrich-Schiller University, Wöllnitzer Str. 7, in Jena. The lake sample was obtained from eutrophic peat lake in North Germany (Lake Zwischenahn) in January 2011 in Meyerhausen, Germany (Zark and Dittmar, 2018). Finally, the Saale river sample was taken in Jena, about 300 m behind the Camsdorfer bridge, ~ 50 cm beneath the water surface against

---

[1] This section has been published as part of the following publication: Simon, C., Roth, V. N., Dittmar, T., & Gleixner, G. (2018). Molecular signals of heterogeneous terrestrial environments identified in dissolved organic matter: a comparative analysis of orbitrap and ion cyclotron resonance mass spectrometers. Front. Earth Sci., 6, 138. doi: 10.3389/feart.2018.00138.

flow, close to the east-ward river bank on 24th of June 2015. About ~ four liters of sample were collected in two combusted Duran glass bottles; the bottles were conditioned three times with sample prior to final collection. Bottles were filled completely. As recommended earlier (Dittmar et al., 2008), all DOM samples were solid phase-extracted using PPL (modified styrene-divinylbenzene polymer) columns, ultrapure methanol (MS grade), ultrapure water and hydrochloric acid (p.a.). Unless when reconstituted from a solid powdery water-free extract, all samples were extracted shortly after sampling.

2.1.2 FT-ICR MS measurements and data processing

FT-ICR MS measurements were performed on a Bruker SolariX equipped with a 15 Tesla cryo-cooled magnet at the Marine Geochemistry group in Oldenburg, Germany (ICBM-MPI-bridging group, Institute for Chemistry and Biology of the Marine Environment, and Max-Planck-Institute for marine Microbiology in Bremen, Germany). This instrument is routinely used for state-of-the-art analyses of complex DOM samples (Hawkes et al., 2016; Riedel and Dittmar, 2014; Zark et al., 2017). All samples were diluted 1:1 (MeOH/ ultrapure water) at a defined dissolved organic carbon (DOC) concentration of 10 ppm (solely the lake sample (BZWA, Table 2-1) was measured at a concentration of 5 ppm) and ionized by electrospray (ESI) in negative mode (Table A-1). SPE-DOM was injected at a flow rate of 4 $\mu l*min^{-1}$. Drying gas temperature was 200 °C. Ion accumulation in a hexapol trap endured for 100 ms. We acquired 500 broadband scans in a range from 115 Da to 2000 Da. To improve the mass accuracy for subsequent formula assignment, raw data were internally recalibrated with a list of ubiquitous formulae (Sleighter et al., 2008) found across environments within the software *Data Analysis 5.0* (Bruker). For this, only peaks with an assignment error ≤ 0.099 ppm were included for recalibration. Data on signal intensities, measured *m/z* and resolution for the whole spectrum were extracted.

Table 2-1. Overview of water samples used in the comparison study. "ID" shows the color code used in graphs. "Set" denotes belonging to full (F) and detailed (D) sets. EE = DOC-based extraction efficiency. References: [1] (Roth et al., 2014); [2] (Roth et al., 2015); [3] (Roth et al., 2013); [4] (Green et al., 2014a); [5] (Zark and Dittmar, 2018); [6] (Schwab et al., 2017); [7] (Green et al., 2014b).

#	Name	ID	System	Site	Set	EE [%]	Ref
1	HS2-5	●	Soil, 5 cm	Deciduous forest, CarboEurope, Hainich, DE	F	72	[1],[2]
2	T-5	●	Soil, 5 cm	Mixed forest, Thann, DE	F	77	[1],[2]
3	W1-5	●	Soil, 5 cm	Conifer forest, CarboEurope, Wetzstein, DE	F	58	[1],[2]
4	JE-2-5-10	●	Soil, 10 cm	Grassland, Jena Experiment, Jena, DE	F, D	61	[1]
5	JE-2-5-20	●	Soil, 20 cm	Grassland, Jena Experiment, Jena, DE	F	76	[1]
6	JE-2-5-30	●	Soil, 30 cm	Grassland, Jena Experiment, Jena, DE	F	69	[1]
7	JE-2-5-60	●	Soil, 60 cm	Grassland, Jena Experiment, Jena, DE	F	64	[1]
8	N3B	●	Bog	Yenisei River Transect, close to Dudinka, RU	F	65	[3]
9	N8B	●	Bog	Yenisei River Transect, close to Bakhta, RU	F, D	65	[3]
10	N8R	●	River	Yenisei River Transect, close to Bakhta, RU	F	65	[3]
11	IHSS	●	River	Suwannee River, S.C. Foster St. Park, GA, USA; Blackwater river	F, D	64	[4]
12	SAALE	●	River	Saale River, Jena, DE	F	52	-
13	BZWA	●	Lake	Zwischenahner Meer, Bad Zwischenahn, DE	F, D	53	[5]
14	H3-2b	●	Aquifer	Oxic Aquifer, Aqua Diva CZO, Hainich NP, DE	F	57	[6]
15	H5-3a	●	Aquifer	Anoxic Aquifer, Aqua Diva CZO, Hainich NP, DE	F, D	59	[6]
16	TAP	●	Tap	Institute of Geosciences, Jena, DE	F, D	37	-
17	NELHA	●	Deep sea	Hawaii, Natural Energy Laboratory of Hawaii Authority, USA. Depth: 674 m	F, D	61	[7]

2.1.3 Orbitrap measurements and data processing

Orbitrap measurements were performed at the Max Planck Institute for Biogeochemistry in Jena, Germany, on a Thermo Orbitrap Elite (Table 1-1). The instrumental settings of the Orbitrap (Table A-1) were optimized in order to yield a reproducible and stable signal that resembled the apparent spectrum shape and the ion abundance distribution of the IHSS reference material (at 20 ppm DOC) on the FT-ICRMS (Table A-2). Eight parameters were evaluated; source fragmentation (SID), inject time and sheath gas were the most influential factors but not checked for interactions. All samples were measured with these optimized instrumental settings. SPE-DOM was injected at a flow rate of 7 $\mu l*min^{-1}$. Drying gas temperature was set to 275 °C. Ion accumulation endured at maximum for 100 ms due to Automatic Gain Control which determines the amount of charge entering the Orbitrap cell. We acquired 500 broadband scans in a range from 115 Da to 2000 Da. The data processing strategy was similar to the data processing in the ICR system and details are given in short below. External mass calibration was performed daily to ensure sufficient mass accuracy for the recalibration and alignment (Liu et al., 2014). Data were obtained by the instrument software LTQ Tune Plus 2.7 (Thermo Fisher Scientific). The data were exported from *Xcalibur* (Thermo

Fisher Scientific) and further processed by the open source software *mmass* (Strohalm et al., 2010), which allows similar processing steps as Bruker's FT-ICR MS software. We transformed spectrum data into a readable format (mzML) using the open software *Proteo Wizard* (Chambers et al., 2012). Peak picking for internal recalibration was performed at 80% peak height in *mmass*. The internal recalibration list was derived from ubiquitous signals found in the FT-ICR MS dataset (Sleighter et al., 2008).

2.1.4 Processing of ultrahigh resolution mass data and comparative analyses

A method detection limit (MDL) was applied to all exported mass lists (Note A–1; Riedel and Dittmar, 2014). The individual spectra were aligned by in-house written software to match similar signals, yielding a crosstab with samples in columns and m/z values in rows. Subsequently, molecular formulae were generated for the matched list of peaks. Formula assignment included the following settings: number of C atoms 1-60; H 4-210; N 0-4; S 0-2; P 0-1; O 1-60, maximal O/C = 1, minimal H/C = 0.3, minimal double bond equivalent = -0.5. These settings relate to assumptions on the probability to encounter specific elements, governed by for example, knowledge on the elemental composition of DOC and solubility, but also known boundaries occupied by organic compounds (Petras et al., 2017). Molecular formulae were only assigned if the mass difference between measured and exact formula mass after internal recalibration was below 0.5 ppm. The probability of multiple formulae assignments per m/z value is increasing with m/z (Koch et al., 2007). The assignment rate (expressed as median per formula) increased to 2 above m/z 600 for both instruments, but increased to 4 above m/z 950 in the Orbitrap data. Formulae with combinations of more than three heteroatoms N, S or P were removed (only N_3 and N_4 being allowed, Rossel et al., 2016). In the following, we only took unambiguous formulae (i.e., single hits) into account. The crosstab was further cleaned for single entries according to published protocols (Mostovaya et al., 2017; Rossel et al., 2016), yielding a final list of formulae. To remove prospective contaminants from the samples, a signal-to-noise ratio (S/N) of blank peaks was defined. The m/z-dependent MDL value at a 99.8 % confidence level was chosen as a noise measure. Only blank peaks with an S/N < 20 were kept in the dataset, accounting for the fact that contaminants often show significantly higher ion abundances. For comparison, all mass spectra were normalized to the sum of their peak intensities (including only peaks > S/N 5). The final crosstabs were handled in two ways for each instrument, depending on whether repeated measurements allowed us to further constrain the information:

1) For a detailed in depth comparison of both instruments, a subset of 7 representative samples covering broad ecosystem groups (bog, soil, river, lake, aquifer, etc.) was measured in replicate on both machines and processed separately (Table 2-1). Only peaks that were detected more than once in all replicates of a sample were included (Riedel and Dittmar, 2014). For Orbitrap data the MDL estimate, originally developed on the ICR instrument, was increased by a factor of 2 to overcome the higher noise level of the instrument (Note A–1, Figure A-1). To evaluate the analytical window of our Orbitrap method, we compared the two groups of formulae that were either detected by both methods and in the same sample ("Common") or only by FT-ICR MS in the respective sample ("FT-ICR MS specific"). The final datasets were used to assess the ion abundance representation, the resolution of heteroatom-containing formulae, and the analytical window of the Orbitrap.

2) For a general comparison of all samples (Table 2-1) including those without the additional constraint of a replicate measurement, we removed formulae that were detected in less than 10% across all measurements to conservatively eliminate noise and treat all samples similar. Replicates remained in the sample set to assess measurement variability but their number was same for all analyses. This dataset was used for assessing biogeochemical information by ordination, i.e. principal coordinate's analysis (PCoA) based on Bray-Curtis dissimilarity and subsequent post-gradient-fitting (*envfit* function of R package *vegan*, at 999 permutations; Osterholz et al., 2016). FTMS formula data of DOM were summarized as weighted averages of chemical indices or a-priori classifications of formulae according to these indices, and applied to constrain the molecular patterns behind separation (Rossel et al., 2016; Roth et al., 2014). A set of indices was calculated for each sample.

2.1.5 Assessment of mass spectrometric resolution

We assessed FT-ICR MS formulae that would be unresolved by the Orbitrap, following the approach of Hawkes et al. (2016). For this, we used a series of triplet signals of $[CHO]N_2O_2$, $[CHO]C_5$, and $[CHO]H_4S$ that are known for their small mass differences. These closely-spaced triplets are distributed over the whole m/z range and provide basis to estimate the actual resolving power of the Orbitrap. In this analysis we used only full triplets, i.e. the $[CHO]N_2O_2$ formula was followed by the other two formulae. $[CHO]N_2O_2$ formulae that were found alone or with only the accompanying $[CHO]C_5$ formula and single $[CHO]C_5$ and $[CHO]H_4S$

formulae were not taken into account. The resolution of the MS can be estimated from the mass differences between these signals at half intensity applying the full width at half maximum (FWHM) criterion of a peak. FWHM data were exported from *mmass* and from the six different sample types a general equation to calculate the FWHM was derived:

$$FWHM = 2 * 10^{-7} * m/z^{1.52} \qquad (1)$$

Based on this is the mass difference between two peaks that can be resolved is 4.1 mDa at ~ 425 m/z and 3.2 mDa at ~ 361 m/z, indicating that up to these m/z the mass differences between the triplet peaks $[CHO]N_2O_2$, $[CHO]C_5$, and $[CHO]H_4S$, respectively, can be theoretically resolved in the Orbitrap system.

2.2 Whitesands and their link to blackwater DOM evolution in the Rio Negro basin

2.2.1 Field sites and sampling procedures

Soil water samples were taken in early November 2017 at two protected forest reserves under responsibility of the Instituto Nacional de Pesquisas da Amazônia (INPA) in Manaus, Brazil (Figure 2-1; Figure 2-2). Both reserves, the Reserva Biológica do Cuieiras – ZF2 (2°36'32.67" S, 60°12'33.48" W, at 40–110 m above sea level) and the Reserva Biológica de Campina (2°35'30.26" S, 60°01'48.79" W, at 93–101 m a.s.l.) are located about 60 - 70 km north of Manaus (De Oliveira Marques et al., 2016; Zanchi et al., 2014). Sampling was conducted from 31st October – 2nd of November, and all lab procedures were conducted within the following three days. The geological setting, landscape structure, forest composition and soil characteristics are described in detail by Zanchi et al. (2014) and references within. In short, the Reserva Cuieiras is characterized by a typical mixture of flat plateaus (26% of the area), broad swampy valleys (43%) and moderately steep slopes in between (31%, total area 22.735 ha, Zanchi et al., 2014). The valley soils show marked difference to the clayey plateau (Oxisols) and slope soils (Ultisols): They are characterized by high sand content (mostly bleached quartz; Podzols, Gleysols), show high content of phenolics and are poorly drained (De Oliveira Marques et al., 2016; Monteiro et al., 2014; Zanchi et al., 2014). Additionally, these soils can store significant amounts of organic matter (~150.000 Mg C*ha^{-1} in the first 120 cm) despite their high macroporosity (De Oliveira Marques et al., 2016). The area is drained by a second-order blackwater stream (Rio Aşu) that meets with Rios Cuieiras and Branquinho (SW of the site) to ultimately drain into the Rio Negro (Monteiro et al., 2014, Figure 2-1). The catchment area is estimated by 660 ha (Monteiro et al., 2014). Monteiro and coworkers report in total

annual rainfall of 2806 mm in 2002 and 2004 mm in 2003. Stream discharge was 0.08 – 5.59 $m^3 s^{-1}$ (average, 0.18 $m^3 s^{-1}$) and DOC levels ranged from 3.2 – 15.2 mg L^{-1} (average, 8.7 ± 3.0 mg L^{-1}) during that period. The annual total carbon flux by the watershed was estimated as 13.3 g m^{-2} in 2003 (Waterloo et al., 2006) and Monteiro and coworkers estimated a stream flux (Rio Aşu) of organic carbon (DOC) equal to 8.7 g m^{-2} for the same year. Zanchi et al. (2015) estimate the annual DOC export of the Aşu watershed to lie within the range 9.3 – 22.7 g m^{-2}.
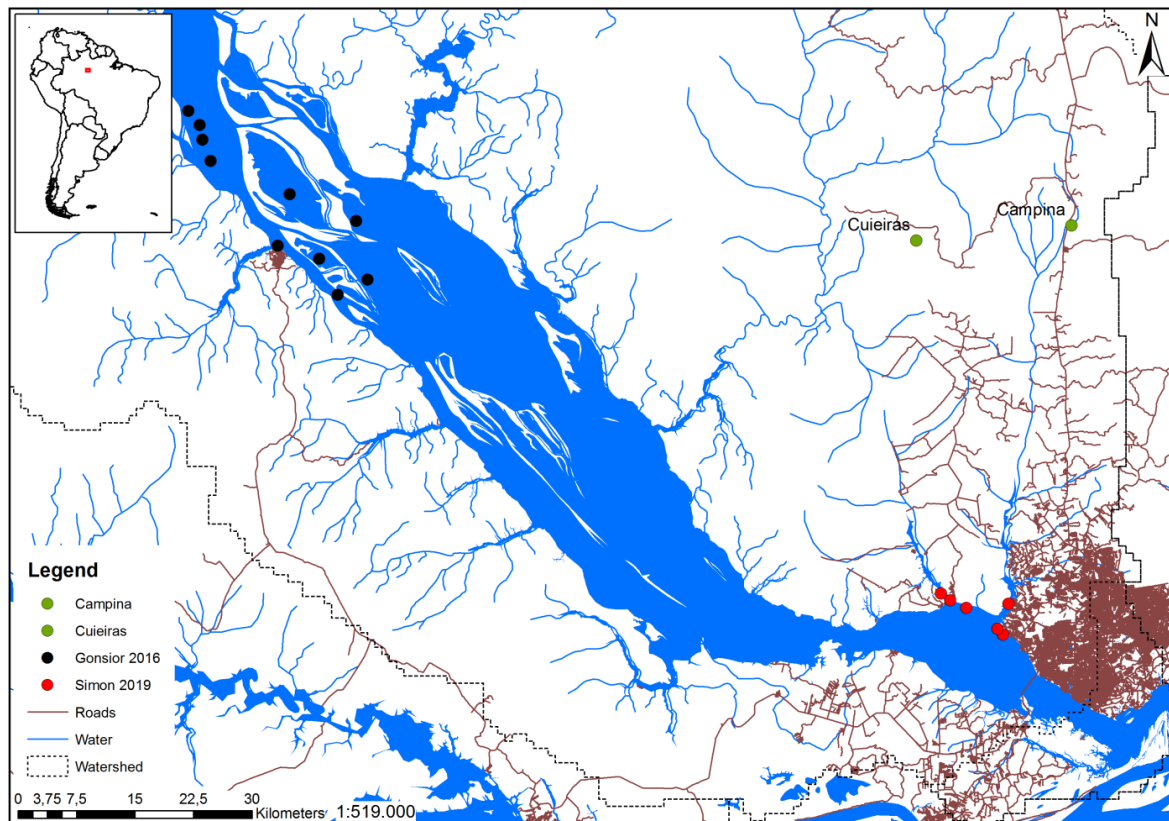
Figure 2-1. Detail of the lower Rio Negro catchment northwest of Manaus. The map shows sampling location of recent FTMS studies. The two sampling locations north of Manaus are marked by green circles ("Cuieiras", "Campina"). For comparison, we also used other FTMS data available online from blackwater samples, including the Rio Negro and connected lakes (black circles; Gonsior et al., 2016) and the lower reach of the Rio Negro and two of its tributaries (Rios Tãruma Mirim and Tãruma Aşu) close to Manaus (red circles; Simon et al., 2019). Watershed limits (obtained by semi-automatic extraction from DEM GTOPO30) and river data (digitalized from a JERS-1 SAR mosaic, with an average density of about 0.15 km/ km²) were accessed as shapefiles from www.ore-hybam.org (Seyler et al., 2009). Headwater stream width is not drawn to scale. Roads and main water bodies were extracted from the OpenStreetMap project (OSM; natural features and roads) and downloaded as shapefiles from www.download.geofabrik.de. Map editor: Marcus Guderle, MPI Jena.
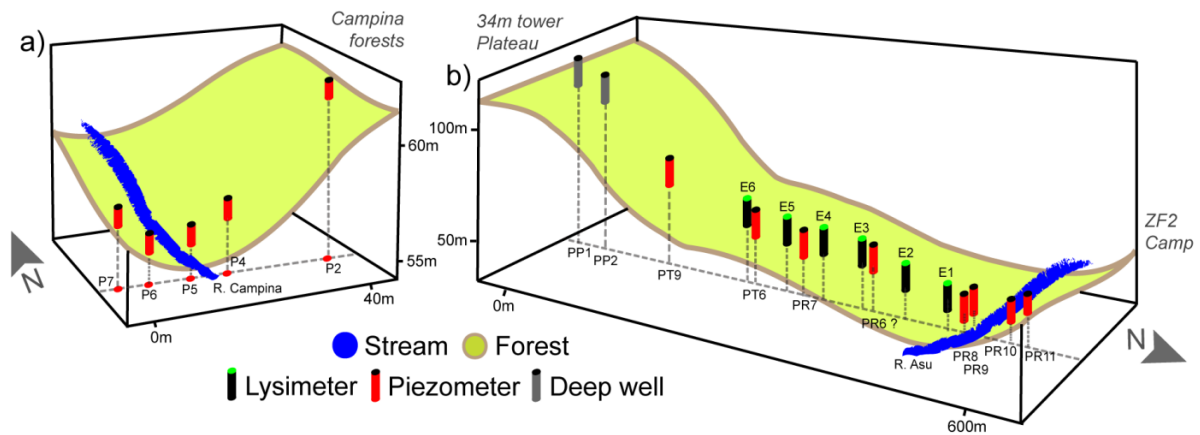
Figure 2-2. Schematic landscape sections of the two sampled ecosystem transects. a) Reserva Campina site (McClain et al., 1997), b) Reserva Cuieiras site (Monteiro et al., 2014; Waterloo et al., 2006; Zanchi et al., 2014). Note differences in scale.

Water samples were taken from lysimeters and piezometers installed across an Aşu valley transect (Monteiro et al., 2014). Lysimeters and piezometers are maintained and sampled at a regular basis. Lysimeters were left at 600 mbar for 2 hours. Piezometers were emptied once before final sampling. The stream was sampled manually, with nitrile gloves, against the direction of flow. Pre-cleaned polycarbonate bottles (acidified ultrapure water (pH2, HCl), square bottles, Nalgene) were cleaned with the respective sample before final sampling. Two deep wells on the plateau (35 m and 39 m depth) were sampled by lowering an empty, clean sampling bottle on a string until water was reached. The Reserva Campina is a smaller reserve (900 ha) and shows much less relief; the area is characterized by extremely poor sandy soils (up to 99% sand) that co-occur with a typical but very specific forest types, so called *Campina* and *Campinarana* forests (heath forests; Demarchi et al., 2018; Targhetta et al., 2015). In contrast to typical highly diverse plateau forests (*terra firme* forests), Campina forests show extraordinary low number of species and have a small canopy height (~10 m vs. 25-40 m at Reserva Cuieiras). Bare patches of sand cover ~11% of the area (Zanchi et al., 2014). The headwater area is drained by a single blackwater stream that is less than 1m wide and often less than 30 cm deep (McClain et al., 1997) and feeds the Rio Taruma Açu in southward direction, meeting with the Rio Negro close by Manaus (Figure 2-1). The catchment area is estimated by 6.5 ha (Zanchi et al., 2015). Zanchi and coworkers report annual rainfall for the period 18th March 2007 until 18th March 2008 (3054mm) which was an exceptionally wet La Ninã year, but stating that rainfall was comparably high during the period in the nearby Cuieiras site. Surface runoff (from the stream) and estimated groundwater outflow amounted to 485 and 1071 mm (1556 combined; as compared to Aşu 1362 mm, Waterloo et al., 2006), respectively, for

that period, with a total amount of 49.2 g m^{-2} (DOC) exported by the watershed (including groundwater outflow and rainfall) of which 15.3 g m^{-2} were exported by the stream. An annual streamflow of 485 mm equals an estimated discharge of 1.0 x 10^{-3} m^{3}s^{-1} which is about 180 times lower than average discharge at Rio Aşu in 2002 – 2003 (Monteiro et al., 2014) but catchment size differs by a factor ~100 as well. McClain and coworkers reported annual DOC exports of 40 g m^{-2} (Zanchi et al., 2015; McClain et al. 1997) from the same catchment in the time period 1993 – 1994. Samples were taken at the side slopes of the stream from piezometers and lysimeters installed in 1993 as detailed in McClain et al. (1997). For this purpose, wells were emptied three times and sampled thereafter (Zanchi et al., 2015). Maximal vacuum applied was 600 mbar. The stream was sampled as described above.

2.2.2 Water chemistry

Aliquots of the samples were subjected to TOC analysis in the water laboratory of the Instituto Nacional de Pesquisas da Amazônia (INPA) in Manaus, Brazil (Laboratório de Águas do INPA/ CPRHC – Coordenação de Pesquisas em Recursos Hídricos e Clima). Samples were measured on a total organic carbon analyzer (TOC-V$_{CPH}$ model, Shimadzu, Kyoto, Japan) as detailed elsewhere (De Oliveira Marques et al., 2012; Monteiro et al., 2014). In short, inorganic carbon is purged from the sample after acidification with phosphoric acid and DOC is determined as NPOC (non-purgeable organic carbon). Persulfate oxidation under heat and UV illumination converts NPOC to CO_2 which is detected through absorption at infrared wavelengths. Prior to extraction, samples were also analyzed for pH and electrical conductivity (EC) with a Multi 340i probe system (WTW, Weilheim, Germany).

2.2.3 Solid-phase extraction

DOM samples were solid phase-extracted (SPE) shortly after sampling in the laboratory of aquatic ecosystems of the INPA, Manaus (Laboratório de Ecossistemas Aquáticos) using an established protocol (Dittmar et al., 2008). The solid-phase sorbent was a modified styrene-divinylbenzene polymer (PPL Bond Elut™, Agilent, Santa Clara, CA, USA). Samples were acidified to pH 2 with 37% hydrochloric acid (Merck EMSURE®, p.a., ACS grade) before extraction. Solvents used for extraction were ultrapure water, acidified ultrapure water (pH2, HCl) and ultrapure methanol (Biotec Reagentes Analíticos, p.a., ACS grade). Columns were loaded with maximal amounts of 3 mg-C based on DOC data. DOC was estimated for Campina samples with a linear regression based on electrical conductivity of Cuieiras piezometer

samples, as demonstrated in Monteiro et al. (2014). Later TOC measurements of these samples at the MPI for Biogeochemistry indicated 20% underestimation of DOC concentrations. The extraction efficiency for DOC-rich samples (> 5mg-C L^{-1}) was 76 ± 12% on a carbon basis (n = 14) at a loading ratio of 440 ± 76 (average ± standard deviation; PPL: DOC in mg/ mg) with a tendency to lower efficiencies at Reserva Campina. For samples low in DOC (n = 9), loading ratios were higher due to sampling constraints (on average, 3000, samples, and extraction efficiencies were either very low (< 30 %; PR11, PT09, PP1, PP2) or too high (> 90%; PR06, E3, E5, E6).

2.2.4 Water isotopes

Isotopic signatures of water (δ^2H and δ^{18}O) were analyzed by high-temperature conversion-isotope ratio mass spectrometry (HTC-IRMS) in the stable isotope laboratory of the Max Planck Institute for Biogeochemistry (BGC-IsoLab). For method details we refer the reader to Gehre et al. (2004). The measurements were conducted on a Delta+ XL coupled to a high temperature furnace via a ConFlow III interface (Thermo Fisher Scientific, Bremen, Germany). One μl of water was injected using an A200S autosampler (CTC Analytics AG, Zwingen, Switzerland) and the furnace temperature was held at 1350 °C. The δ^2H and δ^{18}O values are reported on the VSMOW-SLAP scale which is realized by parallel analysis of samples against in-house standards. In-house standards are routinely calibrated against internationally accepted water standards. Daily standard deviations for δ^2H and δ^{18}O measurements are usually better than 1 and 0.1 ‰, respectively. For comparison, we extracted water isotopy values in rain for October and November and both sampling sites with the help of the online isotopes in precipitation calculator (OIPC2.2, version 3.1, http://wateriso.utah.edu/waterisotopes/, Bowen and Revenaugh, 2003). The values were similar for both sites; -7 ‰ and -17 ‰ (V-SMOW) for δ^2H, and -2.4‰ and -3.8‰ (V-SMOW) for δ^{18}O in October and November, respectively. We estimated the average between both monthly values for each isotope by considering our sampling date at the end of October/ beginning of November.

2.2.5 Radiocarbon analysis

An aliquot of methanolic PPL extract equivalent to 0.25 mg C was transferred to tin capsules (8 mm diameter, 20 mm height; IVA Analysentechnik, Meerbusch, Germany). The methanol was left to evaporate, and caps were refilled until the mentioned amount of C was reached. The air-dried capsules were then combusted in an elemental analyzer, and graphitized for

radiocarbon analysis on a 3 MV Tandetron ^{14}C-AMS (HVEE, Amersfoort, Netherlands) at the Max Planck Institute for Biogeochemistry in Jena, Germany, along with modern (Oxalic Acid II) and ^{14}C-depleted standard materials (Benk et al., 2018; Steinhof et al., 2017, 2010). The graphitization is conducted with Duran glass tubes at a temperature of 550°C in the presence of hydrogen and with an iron catalyst (Steinhof et al., 2017). Radiocarbon concentrations are given as fraction modern (F14C), which is the fraction of the standard concentration including normalization for $\delta^{13}C$ of the Oxalic acid standard measured at the same time (with $\delta^{13}C$ of -19‰ (Trumbore et al., 2016). We then calibrated calendar age estimates from background corrected F14C values. $\Delta^{14}C$ values also take into account radioactive decay of the oxalic acid standard since 1950, which demarks the start of bomb test that increased the ^{14}C content of the atmosphere (Trumbore et al., 2016).

Initially high atmospheric radiocarbon values (due to the "bomb" peak) are constantly diluted by combustion of fossil fuel and thus F14C values may not be ambiguous anymore when "young" samples are measured (Graven, 2015). We thus calibrated our data in two ways: a) as intended, online with "Bomb 13 SH3" data to account for modern F14C values and sampling locations within the inter tropical convergence zone (ITCZ) using *OxCal* online 4.3 (https://c14.arch.ox.ac.uk/oxcal.html, Bronk Ramsey, 2009; Hua et al., 2013), and b) manually with data provided by Graven et al. (2015) for the same geographical zone. For b), we first predicted $\Delta^{14}C$ values for the time period after 2015 based on data from 1997-2015 (linear regression, $R^2 = 0.996$). We then estimated mean calendar age based on this updated record of atmospheric radiocarbon values. The measurement error was < 1 year and < 2 years for repeated measurements (standard deviation). Calendar ages are just an approximation presuming no exchange of initial C and should thus not be interpreted as a definite age. In the case of highly mobile DOC which is in constant exchange with minerals, soil organic matter, plant leachate, root exudates, and soil fauna, this assumption may not be valid. Radiocarbon then becomes a measure of exchange with recently-fixed C. Apparent young ages of DOM can be related to stronger exchange of the DOC pool with recently fixed atmospheric C, i.e., through increased productivity or exudation.

2.2.6 Orbitrap measurements and data processing

Orbitrap measurements were conducted at the Max Planck Institute for Biogeochemistry as described in chapter Orbitrap measurements and data processing, with the only exception that

resolution was set to 480.000k, an option only available through the instrument's development kit (Thermo Fisher Scientific, Zhurov et al., 2013). Allowable numbers of atoms in formula calculation were as follows: $^{12}C_{0-60}$; $^{1}H_{0-120}$; $^{14}N_{0-2}$; $^{32}S_{0-1}$; $^{16}O_{1-60}$, $^{13}C_{0-1}$, and assignment was done at ± 1ppm tolerance. Noise was defined as the smallest peak in the sample set without zeros. We then calculated the maximal signal-to-noise ratio (S/N$_{max}$) of a peak across all samples (without blanks and Saale reference) and only kept peaks with a value > 5. We calculated the number of matches of each peak across all samples (without blanks and Saale reference) and only kept peaks with more than one match. To exclude contaminants, we calculated a signal-to-noise ratio of blank peaks (only blank samples) and discarded all peaks with values > 20. Finally, we excluded those peaks that were only present in less than 20% of all samples (i.e., < 9 matches) and small (SN$_{max}$ < 20). Peaks outside of the usual mass defect range of natural organic compounds were excluded from further analysis (0.3 − 0.95 mDa). 13658 peaks remained in the dataset in total; of those, 9893 had a formula assigned (3672 incl. one ^{13}C atom). Formulae having H/C ratios > 2 or DBE-O values (double bond equivalent minus oxygen atoms) > 15 (modulus) or containing the elemental combination N_2S were discarded due to their unlikely occurrence in DOM (Hawkes et al., 2016; Rossel et al., 2016). In case of ambiguous peaks with more than one assigned molecular formula, only suggested CHO formulae with a DBE-O of < 10 (modulus) were kept. Other formula suggestions or ambiguous hits were excluded from the formula pool and kept as "no reference" peaks. ^{13}C-containing formulae that were missing their equivalent monoisotopic formula were treated similar. As a last measure, we only considered peaks detected twice in two separate runs for further analysis (Riedel and Dittmar, 2014). The final dataset contained 7705 formulae (of those, 1963 containing a ^{13}C). For comparison of samples, all mass spectra were normalized to the sum of their peak intensities (including all peaks > S/N 5, also those with no assigned formula). The further analysis of the data focused on the subset of peaks with an assigned monoisotopic formula (n = 5742).

2.2.7 Statistical analysis

The molecular formula data was analyzed by Principal Coordinate Analysis (PCoA, *cmdscale* function, stats package, v3.5.1) and post-ordination gradient fitting analysis (*envfit* function, vegan package, v2.5-2) within the statistical computation environment R Studio (v1.1.453, © 2009-2018 RStudio, Inc.). PCoA was based on Bray-Curtis dissimilarities obtained by the function *vegdist* from package vegan (see above). The method allows the comparison of

samples based on their formula populations. Redundancy within the dataset, i.e., formulae showing similar trends in ion abundance across samples, is effectively reduced and yields a set of coordinates that summarize the variability of the dataset best. The coordinates are ordered according to their explanatory power, and samples are located in coordinate space according to their coordinate score. The distribution of samples in coordinate space is then analyzed for clustering (indicating similarity among samples) and correlations with specific sample properties with *envfit* (at 999 permutations). We assessed variables from water chemistry analyses, but also derived new variables that statistically summarize properties of the molecular formula population of each individual sample. These variables are shown in summary in Table A-3 and are discussed in further detail in other references (Rossel et al., 2016, 2013; Seidel et al., 2014).

We extracted the subsets of molecular formulae that were significantly correlated with values of Fraction Modern (F14C; Pearson's r, p<0.01) and apparent calibrated calendar ages. Positive correlations with F14C, i.e., older mean calendar ages, suggest a lower exchange of these molecular formulae with recently-fixed carbon whereas negative correlations suggest the opposite, i.e., younger mean calendar ages. This is due to declining atmospheric ^{14}C levels by dilution with radiocarbon-dead fossil fuel CO_2 (Graven, 2015). Assuming similar decay behavior for all ^{14}C atoms, higher values of remaining radiocarbon (F14C) in DOM do thus relate to pools of carbon that were fixed from atmospheric $^{14}CO_2$ at an earlier time point. We also extracted formulae related to the different clusters of samples as derived from PCoA. We conducted two-sided Student's t-tests (assuming unequal variances) based on the relative ion abundances of formulae across samples of each cluster and extracted formulae with significantly (p = 0.05) higher ion abundance for visualization and further comparison.

Subsets of site-specific formulae were compared to molecular formulae indicative of Rio-Negro DOM as derived from a previous study (Simon et al., 2019) and another open-access FTMS dataset (Gonsior et al., 2016). In the previous study, differences of river waters from the confluence of the Amazon near Manaus (Encontro das Águas, "meeting of the waters") were analyzed including three blackwater rivers (Rio Negro, Rio Tarumã Açu, Rio Tarumã Mirim). Gonsior and coworkers compared samples from Rio Negro and some of its adjacent lakes (close to Novo Airão, 120 km northwest of Manaus, Figure 2-1) to river samples from Rio Tapajos and Rio Madeira to reveal large-scale differences in chemodiversity such as unique DOM signals of the three river basins. Rio Negro-specific formulae were defined slightly

different in our previous study: Three mixing gradients under different conditions (natural setting, lab setting with particles, lab setting without particles > 0.2μm) were considered to extract formulae that showed a significant positive correlation to dominance of Rio Negro vs. Rio Solimões during mixing (irrespective of the fact that they may have been present in Rio Solimões at lower ion abundance as well). For the comparison to headwater DOM presented in this study, we thus extracted those two subsets of formulae that were a) exclusively detected (unique) in Rio Negro samples in the study of Gonsior et al. (2016) and b) correlated significantly to Rio Negro dominance in our previous study and compared both sets individually. The both sets of Rio Negro-specific formulae were the analyzed 1) in a broad manner, taking into account all specific signals available, and 2) in a more conservative manner by using only robust signals. In case of our previous study (Simon et al., 2019), the overlap between specific headwater DOM formulae and those formulae correlated significantly with Rio Negro dominance during the mixing process was done 1) with formulae from one experiment (broad approach; MIX NF, laboratory mixing experiment with non-filtered endmembers, n=721, of those CHNO: 1, CHOS: 29) and 2) with formulae fulfilling this criterion in all three mixing experiments (robust approach, n=299, of those CHNO: 0, CHOS: 1). The unique formulae reported by Gonsior and coworkers (Gonsior et al., 2016) were compared by taking into account 1) all formulae only present in Rio Negro samples and at least detected once (broad approach, n=2619, of those CHNO: 1192, CHOS: 124) and 2) only those signals detected across all 18 measurements of the Rio Negro samples and at a minimum ion abundance of 5E7 (robust approach, n=225, of those CHNO: 4, CHOS:0).

For a general comparison of our whitesand sample set and the full set of individual Rio Negro measurements available to date, we used the whole lists of detected formulae across Rio Negro samples (formulae detected at least once) of both studies, yielding a total of 24 additional blackwater DOM measurements (Gonsior et al. (2016): 18 measurements of ten sampling stations; Simon et al. (2019): six measurements from six sampling stations including two Rio Negro tributaries). We used the data as downloaded. For comparison of the datasets, some small adjustments were needed: Formulae detected below m/z 180 and above m/z 800 were removed from the data. Formulae containing P atoms or two S atoms or three to four N atoms were also removed to account for differences in formula assignment (32 formulae with N_3 excluded in Gonsior et al. 2016, 57 P, two S_2 and twelve N_{3-4} formulae excluded in Simon et al. 2019). The remaining masterlists of each dataset were then merged and compared by Venn diagrams (overlap in terms of formula populations) and individual samples were compared by cluster

analysis. For the latter, the data were compiled to one crosstab and transformed to presence/absence format. The clustering was achieved through combination of function *vegdist* of R package vegan (with Bray-Curtis dissimilarity) and *hclust* of R package stats (with Ward linkage for agglomeration, "ward.d2"). Selected indices (Table A-3) were also compared based on weighted averages of the above named DOM indices (based on spectrum envelope, i.e., ion abundance).

2.3 Identification of ecosystem markers by Orbitrap tandem mass spectrometry

2.3.1 Standard substances and reagents

All compounds (Figure 2-3) were first dissolved in one ml of ultrapure MeOH (amounts given in mg in Table A-4). 1 ml ultrapure water (MQ, 18.2 MΩ*cm @ 25°C, Merck Millipore) was added to each stock and thoroughly mixed. In case of Ellagic acid (**#8**), no complete dissolution could be achieved. 100μl DMSO (Dimethylsulfoxide) were added to the stock solution to aid in dissolution, and vortexed for 15 min at 45°C. Afterwards, the stock solution was centrifuged for 1 minute at 17500 rcf (Hermle Z233 MK-2, Hermle Labortechnik GmbH, Wehingen, Germany). All stocks were immediately frozen (-18 °C) upon further use. Before analysis, the stocks were diluted with pure solvents (50% MeOH abs. in MQ) to a final concentration of 20 - 200 mg-C/l (1 ml final volume, Table A-4) and kept cool (4°C, dark) until analysis. The standard compounds can be grouped according to their structural properties (Figure 2-3): Groups A and B contain only one aromatic ring, and differ in the presence of functional groups (A: mainly carboxyl, B: mainly methoxy). Group C contains larger structures containing at least two ring structures from fused subunits (#**7**, quinic acid and caffeic acid; #**8**, two gallic acid monomers; #**9**, coumaric acid, two gallic acid units, and glucose). Group D contains two flavan-3-ol structures, and group E contains three flavonoids with structurally similar but slightly differing flavon-3-ol structures that were also linked to sugars (glycosides).

2.3.2 Soil DOM Sample

The soil DOM sample was initially taken in early November 2005 from a sintered glass suction plate system installed in 5 cm soil depth at a long-term monitoring site in a ~50-year old spruce (*Picea abies*) forest site at Wetzstein, Germany (50° 27' 13" N, 11° 27' 27" E; Kindler et al., 2010; Roth et al., 2015), and immediately freeze-dried for storage. The DOM sample was reconstituted in acidified ultrapure water (pH 2, hydrochloric acid, p.a.) to a final concentration of ~ 3 mg-C /l and solid phase-extracted (PPL cartridges, modified styrene-divinylbenzene

polymer, BondElut, Agilent, CA, USA) according to a published protocol (Dittmar et al., 2008) at a PPL/ DOC ratio of ~ 1400. SPE-DOM was eluted in MS grade methanol and stored at - 20°C until further analysis. The extraction efficiency was 86.9% ± 1.4% on a carbon basis (arithmetic mean ± standard deviation, n = 3).
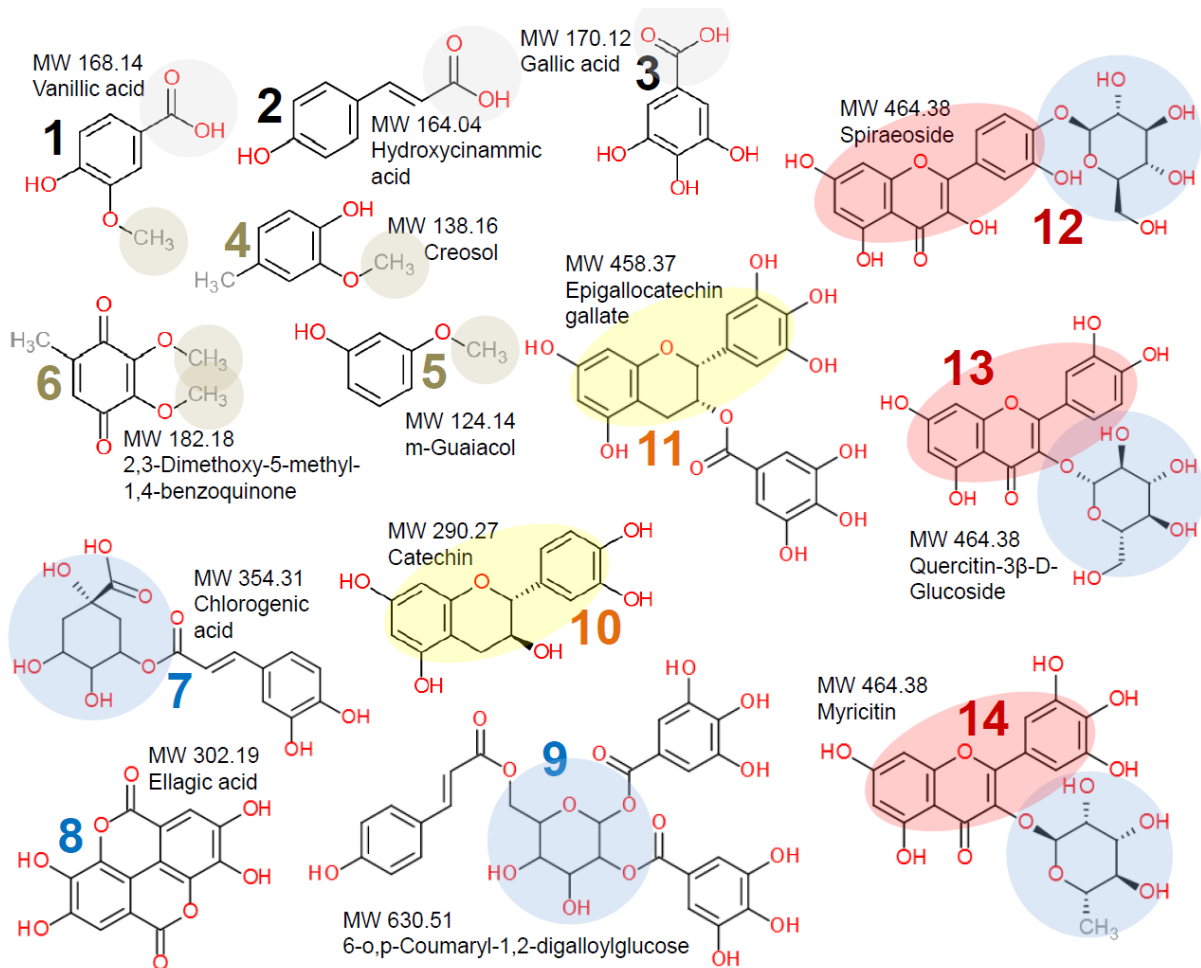
Figure 2-3. Overview of standard compounds used in the study (more information in Table S-2). Colors of the compound IDs refer to the five groups of compound structures analyzed: Group A (black, #1 – #3), Group B (olive, #4 – #6), Group C (blue, #7 – #9), Group D (orange, #10, #11), and Group E (red, #12 – #14). Groups A and B contain only one aromatic ring, and differ in the presence of functional groups (A: mainly carboxyl, grey circles, B: mainly methoxy, brown circles). Group C contains larger structures containing at least two ring structures from fused subunits (#7, quinic acid and caffeic acid; #8, two gallic acid monomers; #9, coumaric acid, two gallic acid units, and glucose). Group D contains two flavan-3-ol structures (yellow oval shapes), and group E contains three flavonoids (red oval shapes) with structurally similar but slightly differing flavon-3-ol structures linked to sugars (glycosides). Polyol units are marked by blue circles.

2.3.3 Orbitrap tandem MS analysis of standard compounds

The standard solutions were infused directly into the ESI (electrospray) source of an Orbitrap Elite (Thermo Fisher Scientific, Bremen) as described in chapter Orbitrap measurements and data processing. The ESI was operated in negative mode and solution was infused at a flow

rate of 10 µl/ min (50% MeOH/ 50% ultrapure water). The Orbitrap response was optimized for each substance by tuning sheath and aux gas flows (Nitrogen), spray voltage, S-Lens RF, the ESI needle position (distance to the instrument inlet), and the scan range that depended on precursor ion mass. The remaining instrument settings were left unchanged for all compounds (settings, Table A-5). We performed MS^2 fragmentation experiments at three normalized collision energy levels (15, 20, 25) in collision-induced dissociation (CID) mode. These collision energies covered the range of near-complete fragmentation of DOM. In some cases, also MS^3 spectra of selected product ions were acquired. Raw data were treated as described in chapter Orbitrap measurements and data processing. We recalibrated the mass spectra by the known exact mass of the precursor ion and plausible product ions at lower *m/z* to improve mass accuracy of unknown product ion peaks, and subsequently, the derived mass differences. Recalibrant ion identity was checked for plausibility by a threefold confirmatory approach: 1) Suggested molecular formula in MIDAS (Formula Calculator v.1.2.6, National High Magnetic Field Laboratory (NHMFL), Tallahassee, USA) based on exact mass and wide elemental constraints; 2) Predicted fragmentation products in Mass Frontier 7.0 (Thermo Fisher Scientific); and 3) Reports of fragment identity (molecular formula and structure) from the literature (recalibrant ions specified in Table A-6). Alignment of fragment mass spectra and molecular formula annotation was achieved via self-written Matlab routines (by Thomas Riedel and Benjamin Jacob). The settings for formula annotation were as follows: Minimum allowed H/C ratio, 0.3; maximum allowed O/C ratio, 1; minimum allowed double bond equivalent (DBE), -0.5; charge, -1; min #C, 1; min #H, 1, min #O, 1. The error of most annotated formulae was within ± 0.5 ppm; the maximum tolerance allowed was ±1 ppm. The upper elemental boundaries for fragment annotation were determined by the standard's neutral molecular formula. Assignments were rechecked with MIDAS, especially the presence of radical anions. All major peaks were annotated with a molecular formula. We removed product ion peaks that occurred only once across the collision energy gradient. We also removed peaks that showed a maximum absolute intensity below 1000 across all fragment spectra. Absolute intensity (ion abundance) was normalized to the intensity of the base peak for each mass spectrum (fragment spectra described in Table A-7). Fragmentation spectra were evaluated with Sirius (Böcker and Dührkop, 2016; Rasche et al., 2012) and CSI Finger ID (Dührkop et al., 2015, Table A-8) for quality control and interpretation of fragmentation spectra.

We calculated the mass differences between the substance's precursor ion (always [M-H]- ions, except substance **#6**, M-•) and all product ions, much similar to a recently described approach

for recalibration of MS^1 data (Smirnov et al., 2019). Separate lists were created for each CID energy mode. In order to exclude unique but less important mass differences from our analysis, we derived a list of those mass differences (n=55) that were 1) either related to a fragment with a minimum relative intensity (basepeak) of 1% or 2) detected more than once across the 14 pure substances (Table A-9). Eight of the mass differences were also described within the DOM literature in DOM, for example the ubiquitous CO_2 and H_2O losses (Hawkes et al., 2018a; Kunenkov et al., 2009; Witt et al., 2009; Zark and Dittmar, 2018), and were thus excluded in some analyses. The comparison of measured mass differences to predicted mass differences (loss from molecular formula) allowed us to assess assignment error of mass differences in our dataset. Above mass differences of 75 m/z, the error between both values was well below 1 ppm (Figure A-2). As expected, the error peaked at values around 5 ppm at a very small mass difference (15 – 30 *m/z*).

2.3.4 Orbitrap tandem MS analysis of DOM

The DOM sample (W1-5) was injected at a concentration of 100 mg-C/l into the above described Orbitrap Elite system. The DOM sample was injected at a five-fold higher carbon concentration than in preliminary studies (Roth et al., 2014; Simon et al., 2018) to compensate for the low concentration of individual compounds and increase sensitivity in tandem MS experiments (Wagner et al., 2015). The instrumental settings to create MS^1 data for precursor ion isolation were similar to the method described before and yielded similar response. All marker signals from the previous study (Roth et al., 2014) were also found by the Orbitrap. The parameters for the MS^2 experiments were same as for the standard compounds if not noted differently (Table A-5). The scan range was adapted to the mother ion mass. All other parameters were left as described in chapter Orbitrap measurements and data processing. Four isobaric precursor ion mixtures (IPIMs; "mass islands", at *m/z* 241, 301, 361 and 417) were chosen for fragmentation because they span the range of maximum ion abundance typically observed in terrestrial DOM samples, and because they each contained one potential tannic forest marker described in the ecosystem study (Roth et al., 2014; the proposed identity of these molecular formulae was only based on their H/C and O/C ratios). The selected potential "tannin" formulae were (monoisotopic masses of $[M-H]^-$ ions are given in brackets): $C_9H_6O_8$ (*m/z* 240.9990), $C_{11}H_{10}O_{10}$ (*m/z* 301.0201), $C_{13}H_{14}O_{12}$ (*m/z* 361.0413) and $C_{15}H_{14}O_{14}$ (*m/z* 417.0311). The set of precursors (n = 159) also contained one potential age marker (neutral formula $C_{16}H_{14}O_6$, *m/z* 301.0717) that showed significant negative rank correlation to

calibrated ^{14}C age in a previous study including the sample (Benk et al., 2018). We collected 150 scans per fragmentation experiment and ran every experiment twice.

The raw data processing followed the same steps as described for standard compounds. Recalibration lists were constructed from known molecular formulae of precursor ions and ubiquitous non-indicative neutral losses (i.e., multiples of CO_2, H_2O, and CO losses, Table A-10; Hawkes et al., 2018; Kunenkov et al., 2009; Witt et al., 2009; Zark and Dittmar, 2018) and applied to improve the mass accuracy of the derived MD data (Smirnov et al., 2019). The final exported peak lists were picked at an absolute signal intensity threshold of 10, equivalent to an S/N > 3. Alignment of fragment mass spectra and molecular formula annotation followed same routines and with similar settings as described for standard compounds except that the elemental boundaries for fragment annotation were: C, 1-40; H, 1-200; N, 0-4; O, 1-40; S, 0-2. For data cleanup, we first removed peaks that were only detected once across all tandem mass spectra as they are prone to be noise. Molecular formulae with unlikely combinations of heteroatoms ($N_{2-4}S$, and $N_{2-4}S_2$) were classified as unassigned peaks, and if multiple formulae were proposed, preference was given to the CHO formula. DOM precursors group naturally in "mass defect islands" (IPIMs) within -0.05 and +0.35 mDa of an integer m/z (Riedel and Dittmar, 2014). The definitions of isobaric complexity and isomeric complexity have to be differentiated in that context: Isobaric complexity evolves from the ultrahigh resolution of the mass spectrometer, while the latter relates to structural isomers that cannot be differentiated based on their exact m/z (and thus, molecular formula; Hertkorn et al., 2013, 2008). Product ions in the mass defect region between 0.35 and 0.95 Da were not observed. Finally, only precursor and product ions detected in both replicate fragmentation experiments were taken into account as an effective measure of excluding false positive signals (Riedel and Dittmar, 2014; Simon et al., 2018).

2.3.5 Mass difference matching procedure

The isolation window of the front end LTQ mass spectrometer in the Orbitrap Elite was set to 1 Da in order not to sacrifice sensitivity and isolate a single IPIM. The ultrahigh resolution and mass accuracy achieved by the Orbitrap allowed us to link individual resolved molecular formulae of precursor and product ions (Osterholz et al., 2015; Zark and Dittmar, 2018). We obtained the MDs of all precursor ions and product ions, yielding a MD spectrum, and combined these into a MD matrix. Each of the four IPIMs and three CID energy levels was

treated individually. The matrix represents all possible combinations of precursor and product ions. To match MDs between the two libraries (indicative MDs and non-indicative MDs) and the MD matrices, exact MDs were kept to the fourth digit. The tolerance for a positive match with the MD spectrum was set to ± 0.0002 Da (2 ppm at 200 Da), which roughly accounts for the mass error from two *m/z* measurements (precursor & product ion). We evaluated the validity of our approach by assessing the probability of a (false positive) match by increasing the tolerance window further (up to 0.002 Da). To analyze patterns of matching frequency we visualized precursor formulae in 3D-Van Krevelen space. The third (z-) axis represents the matching frequency to either whole lists of MDs or subsets of MDs.

2.3.6 Similarity assessment and structure suggestions

We used the information from the matching analysis to compare individual matching profiles of standard compounds and DOM precursors to evaluate potential identities of the underlying unknown structures. We performed this analysis by hierarchical clustering based on Ward's method (using Euclidean distance). The underlying matching data was combined from the standard compound data and IPIM data, separated for the three energy levels and transformed into presence/ absence format. Because highest numbers of matches to the indicative MD list were achieved at CID 25 (Figure A-3), we focused our further analysis on this energy stage. The matching profiles were then compared to lists of structural formula suggestions from the Dictionary of Natural Products (DNP, v27.2, dnp.chemnetbase.com). The DNP is a subset of the Chapman & Hall/ CRC Chemical Database, and references structures of natural products described in the scientific literature.

3 Resolution of molecular detail in DOM by Orbitrap mass spectrometry[2]

3.1 Introduction

Dissolved organic matter (DOM) links the organic matter pools of terrestrial and marine ecosystems through transport of material derived from biota, degrading plant litter and soil organic matter (SOM) to the ocean (Marín-Spiotta et al., 2014; Ward et al., 2017). Taken together, both pools hold about four times more carbon than the atmospheric carbon pool and minor changes in the oxidation or mobilization rates may have major climatic impacts (Carlson and Hansell, 2014; Heimann and Reichstein, 2008). Therefore it is necessary to better understand and identify the underlying processes that control organic matter dynamics. Molecular-level investigations greatly improved our understanding of carbon dynamics in the recent years and highlighted the importance of high resolution analytical techniques in Earth Science (Gleixner, 2013; Kallenbach et al., 2016; Kujawinski, 2011; Marschner et al., 2008; Medeiros et al., 2016; Schmidt et al., 2011; Zark et al., 2017). Ultra-high resolution mass spectrometry like Fourier-transform mass spectrometry (FTMS) coupled to soft ionization techniques as electrospray ionization (ESI, ESI-FTMS) uniquely enables, for example, the identification of thousands of intact (non-fragmented) individual molecular formulae from complex mixtures like DOM (Hertkorn et al., 2013). The observed structural heterogeneity in terms of molecular formulae, which is larger in terrestrial DOM than in marine DOM, encodes source materials, transforming processes and their controlling environmental and biological factors (Bailey et al., 2017; Kellerman et al., 2014; Roth et al., 2014; Seifert et al., 2016; Ward et al., 2017; Zark and Dittmar, 2018). Unfortunately, the advances in our understanding of the molecular DOM "code" are small as the access to ultra-high resolution mass spectrometry is limited.

So far ultra-high resolution analyses have been limited to ion cyclotron resonance (ICR) instruments (Hertkorn et al., 2013; Marshall and Hendrickson, 2008; Qi and O'Connor, 2014) and only few of these systems are available due to the high instrument and maintenance costs

---

of the superconducting magnets needed to achieve ultra-high resolution. The Orbitrap mass analyzer – which uses electrostatic fields for ion trapping instead of magnetic fields (Table 1-1; Zubarev and Makarov, 2013) – is a more economical alternative to FT-ICR MS instruments as it, for example, requires no extensive cooling. However, Orbitrap mass analyzers have a ~10-fold lower mass resolution (at similar transient lengths) that limits separation of peaks in the higher mass range. Although Orbitrap instruments have been successfully applied for characterizing complex natural organic materials (Table A-11) it remains unclear how comparable the results of both instrument types are. Addressing this question is even more pivotal when aiming to compare trends in larger-scale DOM sample sets obtained by different instruments (Swenson et al., 2014).

Previous studies that compared both instrument types used only one or two samples, or their mixtures, for the comparison (Table A-11). The authors underlined the potential of the Orbitrap analyzer and demonstrated general data comparability. While several authors state that the lower resolving power of their Orbitrap instruments limits the application for detailed molecular analysis, they generally proved that fast sample characterization using molecular indicators or fingerprints of the most abundant signals is feasible (Hawkes et al., 2016; Mangal et al., 2016; Pomerantz et al., 2011; Smith et al., 2012). One of the more recent technical developments is the Orbitrap Elite featuring a resolution up to 240000 at m/z 400) (Table 1-1; Denisov et al., 2012; Zhurov et al., 2013). It includes a modified version of the analyzer cell (high-field Orbitrap) and an enhanced FT algorithm ("eFT"), ultimately leading to a significant increase in analytical capacities compared to earlier instruments and reduces the difference in resolution to ICR instruments to a factor of 4. ICR instruments undoubtedly have the overall better analytical performance, especially in the higher mass range. However, for the analysis of the molecular composition of terrestrial DOM, which has an intensity maximum of masses around 400 Da, Orbitrap Elite analyzers might be a cost effective alternative.

The molecular composition of terrestrial DOM, which is a snapshot of the sum of all ecosystem activities, is highly diverse between ecosystems and highly dynamic within ecosystems (Bailey et al., 2017; Gonsior et al., 2016; Hertkorn et al., 2016; Kellerman et al., 2014; Roth et al., 2015, 2014). The reproduction of this complexity has so far not been the focus of previous studies that compared mainly the analytical performance of both analyzers. It is nevertheless necessary to evaluate whether the Orbitrap analyzer resolves the important signals that are needed to separate DOM samples based on their origin and if both analyzers use the same mass

signals to retrieve this information. Similarly, it is also necessary to assess whether the sensitivity of the instruments affects the obtained discrimination, as smaller signals may be undetected by the Orbitrap analyzer.

These questions, which are centered on the information content of the samples, can only be solved using multiple samples that cover a wider range of terrestrial DOM samples. Therefore we analyzed 17 DOM samples from varying environmental settings (Table 2-1) with an ICR analyzer and an improved Orbitrap analyzer and used statistical methods to compare their information content. In detail we assessed the following questions:

a) Does the Orbitrap analyzer detect similar formulae and reproduce the mass abundance patterns of the ICR analyzer in all samples?

b) Are specific masses missing in the abundance patterns of the Orbitrap analyzer, and is this loss related to sample properties, instrumental resolution or sensitivity?

c) Does the Orbitrap analyzer reproduce the discriminating information obtained from FT-ICR MS and from multivariate statistical analyses of the samples, and how well are molecular trends retrieved which were found in previous studies?

3.2 Results

3.2.1 General features

The Orbitrap Elite system expectedly had a higher reproducibility, mass accuracy and resolution than reported for an earlier instrument version (Hawkes et al., 2016), but lower compared to the ICR reference system (Table 3-1). In the Orbitrap the absolute deviation from the exact mass of the molecular formula was by a factor of 6 larger, the interquartile range (IQR) of all mass errors was by a factor of 2 larger and the resolution at m/z 401 was about the factor of 3 lower. The Bray-Curtis dissimilarity between the triplicates was roughly a factor of 2 lower in the Orbitrap indicating that the replicates were more similar in the Orbitrap system compared to the ICR system.

Table 3-1. General properties of both instrument's datasets. IQR, interquartile range; SD, standard deviation.

Property	Orbitrap	FTICRMS
Mass accuracy [ppm] Median (IQR), formulae	-0.06 (0.30)	-0.01 (0.16)
Resolution Mean ± SD, all peaks at m/z 401	215974 ± 1851	615148 ± 13610
Bray-Curtis dissimilarity between triplicates [%] Mean ± SD	2.71 ± 0.76	6.01 ± 1.72

3.2.2 Reproducibility of ion abundance patterns

The normalized total ion abundance patterns proved to be highly similar between both instruments (mass range m/z 115-1000; Table 3-2). More than 90% of cumulative ion abundance were covered by a much narrower mass range (m/z 200-650) in all seven samples, with about 70 – 80% occupied by assigned signals (formulae, Forms). A notable deviation between instrumental response were higher numbers of unassigned signals ("noRefs") in the mass range above m/z 200 for the Orbitrap, especially for the three samples with high input of fresh and freshly degrading plant material (bog, N8B; blackwater river, IHSS; and upper soil, JE-2-5-10).

When only looking at the assigned signals in the mass range m/z 200-650, which forms the basis of the present analysis, the general picture was the same. More than 90% of information (ion abundance) was shared among instruments (Table 3-2, common signals, "Coms"), although the numbers of formulae shared varied broadly from 2687 (H5-3a) to up to 4931 (BZWA). Although similar in number on each instrument, the shared formulae showed more

ion abundance on FT-ICR MS in the mass range m/z 200-400 (up to 12% more, depending on sample), and showed more ion abundance on the Orbitrap in the mass range m/z 400-650 (up to 15% more, depending on sample). In contrast, the ion abundance distribution was remarkably similar in case of the tap water sample (TAP).
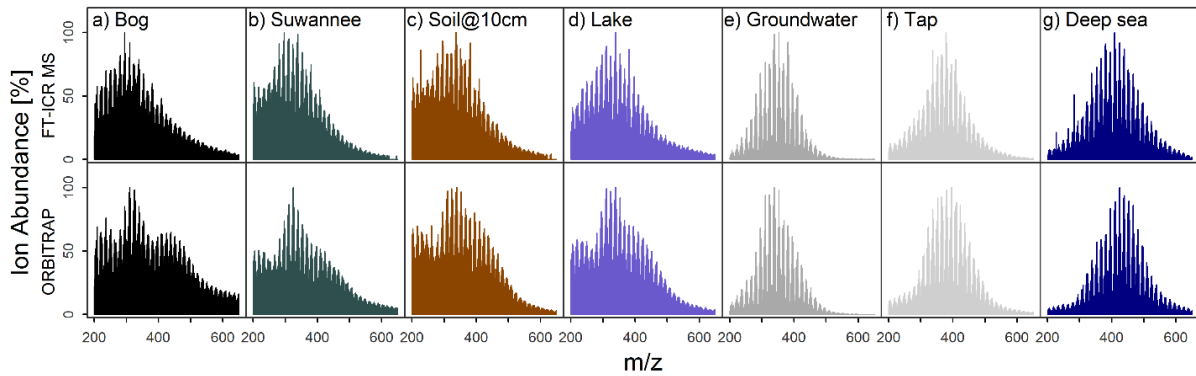
Figure 3-1. Ion abundance patterns of FT-ICR MS and Orbitrap. Seven samples (Table 2–1, set "D") were selected for detailed analyses, showing the general congruence in spectrum shape (only commonly detected signals shown) for samples from a) a bog, b) a blackwater river, c) shallow soil, d) a lake, e) an aquifer, f) tap water and g) the deep sea. Colors are for visual guidance.
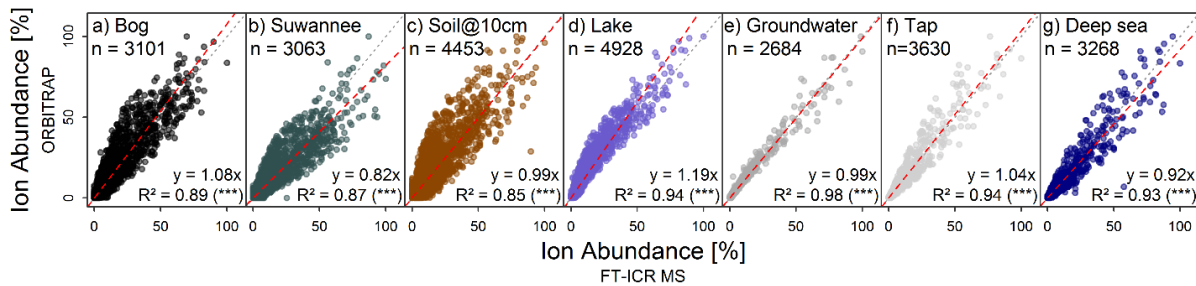
Figure 3-2. Direct comparison of relative ion abundance data of formulae detected by both instruments. Comparison of a) a bog, b) a blackwater river, c) shallow soil, d) a lake, e) an aquifer, f) tap water and g) the deep sea (same as in Figure 1), approved the overall similarity of spectrum shapes at the formula level. Colors are for visual guidance.

Table 3-2. Absolute numbers and relative intensity contribution (in brackets, in % of total) of different groups of signals. Data from seven selected DOM samples measured by Orbitrap (a, b) and FT-ICR MS (c, d). Shown is the comparison between assigned signals (to formulae, Forms) and unassigned signals (noRefs) across the mass range m/z 115 – 1000 (a, c), and between instruments-shared (common, Coms) and instrument-specific (Specs) assigned signals (only formulae) over the range m/z 200-650 (b, d).

Sample	a) Orbitrap: m/z 115-1000, all signals				b) Orbitrap: m/z 200-650, only formulae (Coms + Specs)					
	NoRefs < m/z 200	Forms m/z 200 – 650	NoRefs m/z 200 – 650	NoRefs > m/z 650	Coms total	Coms < m/z 400	Coms > m/z 400	Specs total	Specs < m/z 400	Specs > m/z 400
N8B	809 (8.7)	5276 (73.9)	5423 (12.7)	2302 (4.7)	3104 (97.7)	1581 (54.1)	1523 (43.6)	2172 (2.3)	805 (0.6)	1367 (1.7)
IHSS	1109 (10.7)	5899 (73.6)	6369 (13.9)	2110 (1.8)	3066 (95.4)	1817 (59.7)	1249 (35.7)	2833 (4.6)	1022 (0.9)	1811 (3.7)
JE-2-5-10	1240 (12.1)	7057 (74.7)	6660 (12.5)	951 (0.7)	4456 (95.2)	2757 (63.5)	1699 (31.7)	2601 (4.8)	757 (0.8)	1844 (4.0)
BZWA	1093 (11.4)	5977 (74.7)	5659 (13.1)	917 (0.8)	4931 (99.0)	2491 (63.2)	2440 (35.9)	1046 (1.0)	497 (0.6)	549 (0.4)
H5-3a	492 (2.9)	3438 (83.3)	2737 (13.5)	523 (0.3)	2687 (99.4)	1591 (77.4)	1096 (22.0)	751 (0.6)	386 (0.2)	365 (0.4)
TAP	774 (3.2)	4867 (80.8)	4447 (14.9)	1237 (1.1)	3633 (99.5)	1883 (55.1)	1750 (44.4)	1234 (0.5)	588 (0.2)	646 (0.3)
NELHA	415 (1.2)	4290 (78.6)	3564 (18.5)	1683 (1.7)	3271 (98.6)	1582 (34.0)	1689 (64.6)	1019 (1.4)	408 (0.2)	611 (1.2)
Sample	c) FT-ICR MS: m/z 115-1000, all signals				d) FT-ICR MS: m/z 200-650, only formulae (Coms + Specs)					
	NoRefs < m/z 200	Forms m/z 200 – 650	NoRefs m/z 200 – 650	NoRefs > m/z 650	Coms total	Coms < m/z 400	Coms > m/z 400	Specs total	Specs < m/z 400	Specs > m/z 400
N8B	743 (10.4)	3597 (70.1)	2705 (18.8)	162 (0.7)	3104 (96.6)	1581 (65.6)	1523 (31.0)	493 (3.4)	240 (1.9)	253 (1.5)
IHSS	846 (12.6)	3545 (67.9)	2655 (19.4)	11 (0.1)	3066 (94.7)	1817 (71.2)	1249 (23.5)	479 (5.3)	302 (3.7)	177 (1.6)
JE-2-5-10	1096 (13.3)	5744 (71.2)	2499 (15.5)	7 (0.0)	4456 (89.7)	2757 (68.0)	1699 (21.7)	1288 (10.3)	861 (7.6)	427 (2.7)
BZWA	1105 (7.7)	6822 (72.1)	4747 (17.9)	840 (2.3)	4931 (93.3)	2491 (61.3)	2440 (32.0)	1891 (6.7)	668 (2.5)	1223 (4.2)
H5-3a	586 (2.6)	3722 (81.7)	2504 (15.6)	96 (0.1)	2687 (96.7)	1591 (75.1)	1096 (21.6)	1035 (3.3)	377 (1.3)	658 (2.0)
TAP	747 (4.2)	4653 (76.4)	3353 (19.2)	103 (0.2)	3633 (96.0)	1883 (58.6)	1750 (37.4)	1020 (4.0)	412 (1.7)	608 (2.3)
NELHA	774 (3.3)	5219 (76.2)	3534 (19.3)	385 (1.2)	3271 (89.2)	1582 (39.9)	1689 (49.3)	1948 (10.8)	443 (2.5)	1505 (8.3)

Regarding specific formulae ("Specs", Table 3-2), numbers were higher in the Orbitrap but cumulative ion abundance was higher in the FT-ICR MS data, being somewhat congruent to the observation made for the whole mass range. Again, this was especially pronounced in the three samples with high inputs of fresh and freshly degrading plant material (N8B; IHSS; JE-2-5-10, Figure A-1). Samples BZWA, H5-3a, TAP, and NELHA showed higher numbers of FT-ICR MS specific formulae detected above m/z 400, going along with a higher ion abundance compared to number and ion abundance of specific formulae detected below m/z 400. The shape of normalized total ion abundance patterns were in good agreement between both instruments (Figure 3-1, Figure 3-2). This analysis considered only formulae detected by both instruments, which also constituted the majority of ion abundance on both instruments, as explained above. Whereas the deep sea (NELHA), aquifer (H5-3a) and tap (TAP) water samples had almost identical patterns, lake (BZWA), bog (N8B), soil (JE-2-5-10) and Suwannee river (IHSS) samples showed slight deviations in the m/z range centered at m/z ~ 225 and at m/z ~ 450, which was also reflected by a higher mean Bray-Curtis dissimilarity among instruments (13.9 ± 3.5 vs. 19.9 ± 1.7). The shared molecular formulae, which represented the majority of detected signals as measured by their contribution to the overall ion abundance, agreed also very well (Figure 3-2, Figure A-1). For the aquifer, deep sea, tap water and lake samples R^2 of the regression line was larger than 0.9. For samples that were more directly influenced by inputs of fresh and initially decomposing plant litter and thus also show higher overall DOC concentrations (IHSS, N8B, and JE-2-5-10) R^2 varied between 0.85 and 0.89. The slope coefficients obtained for most samples were within 10% range of the 1:1-line, which corresponded to the slope variability (per instrument) for replicated measurements. In contrast, the slope coefficients of the remaining three samples (BZWA, IHSS and NELHA) deviated stronger from their replicate slopes (1.2, 0.8, 0.9). A general feature observed on both instruments was the apparent increase in average mass to charge ratio in the order: Bog ~ Suwannee river ~ grassland soil (10cm) < lake water < aquifer ~ tap water < deep sea.
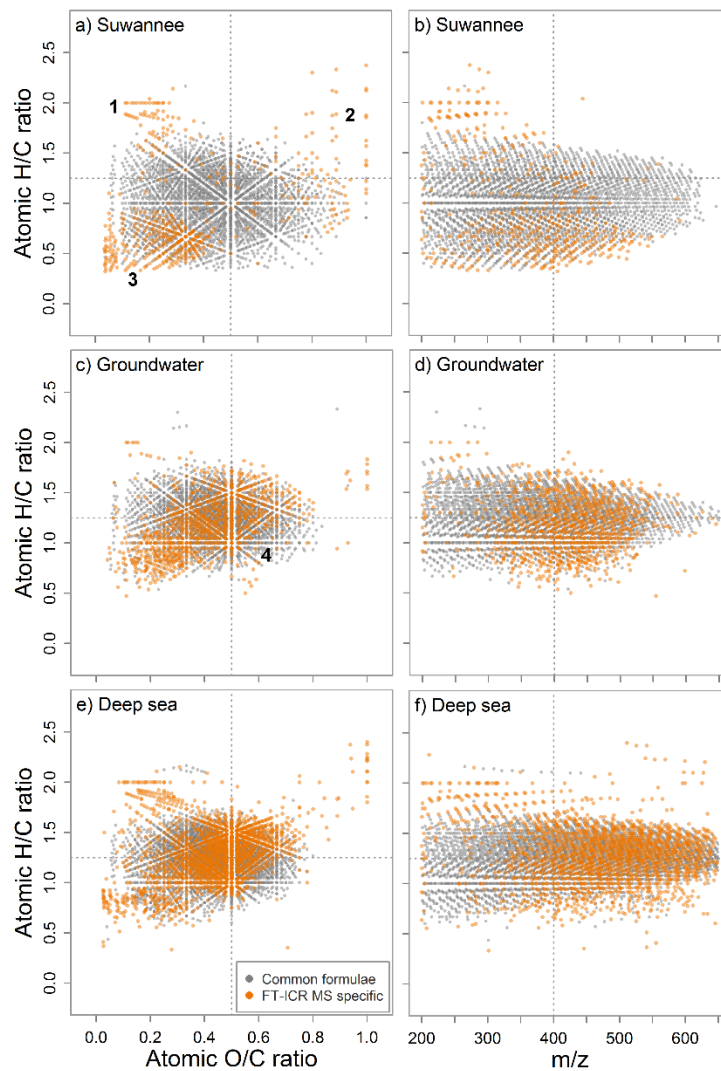
Figure 3-3. Comparison of common (grey) and FT-ICR MS-specific formulae (orange), by their van-Krevelen fingerprints (a, c and e) and saturation state over mass (H/C vs. m/z, b, d and f), here shown for three samples spanning a heteroatom gradient from the Suwannee river (IHSS) over the aquifer (H5-3a) to the deep sea (NELHA) sample. Four dedicated formula clusters (1 – 4) were found to be additionally detected by FT-ICR MS.

3.2.3 Analytical window: Assessment of information loss

The IHSS sample was characterized by lowest contribution of heteroatom formulae including N, S and P atoms within the subset of samples used for detailed analyses (mean molecular formula: $C_{17.3}H_{18.2}O_{7.4}N_{0.1}S_{0.05}$; 38.7 % NSP-containing formulae), whereas the samples from deep sea ($C_{20.1}H_{25.8}O_{8.9}N_{0.4}S_{0.04}P_{0.02}$; 61.0%) and an anoxic aquifer (H5-3a; $C_{17.4}H_{22.0}O_{7.9}N_{0.2}S_{0.02}P_{0.01}$; 61.3% NSP-containing formulae) showed highest contributions marking the "endmembers" of elemental diversity in the detailed sample set. The number of formulae that were specific for FT-ICR MS was lower in the IHSS sample (n = 497, ~ 13%) than in the other samples (H5-3a, n = 1035, ~ 28%; NELHA, n = 1948, ~37%) and the specific formulae were differently distributed in van-Krevelen space (Figure 3-3). Three distinct

clusters of FT-ICR MS-specific formulae were found in van-Krevelen space of the IHSS sample (Figure 3-3a): 1) The very saturated (high H/C, low O/C) region commonly referred to as lipid- or peptide-like substances (D'Andrilli et al., 2015). 2) The highly oxygenated formula region (high O/C, variable H/C) often termed as "carbohydrate-like" substances (D'Andrilli et al., 2015). 3.) The region of formulae relatively poor in both H and O (low H/C, low O/C) that is linked to highly (poly-) aromatic, phenolic or black carbon-like substances (Figure 3-3a; Koch and Dittmar, 2016). These clusters differed also in their mass distribution patterns (Figure 3-3b). While members of cluster 1 reached molecular weights of up to m/z 350, the rather moderately saturated second cluster reached up to m/z 500. The hydrogen depleted third cluster spread over the whole mass range with a center at m/z 400.

In the heteroatom-rich samples (Figure 3-3c and e) a fourth formula cluster of FT-ICR MS-specific signals was observed in the region commonly assigned to carboxyl-rich alicyclic molecules (CRAM, Hertkorn et al., 2006) and materials derived from linear terpenoids (MDLT, Lam et al., 2007). In contrast to the black-water river (IHSS) sample, we found clear differences in contribution of the first three clusters to the overall pool of FT-ICR MS-specific formulae, being linked to a shift in the center of ion abundance distribution. The anoxic aquifer (H5-3a, Figure 3-3d) was characterized by a DOM ensemble showing only minor contribution of the three first groups while the deep sea sample (NELHA, Figure 3-3e, f) was very rich in all of them, as compared to the IHSS sample. High molecular weight carbohydrate like-compounds and elongated series of higher-molecular weight lipid- and peptide-like formulae were detected additionally, along with unsaturated formulae showing the same shift to higher *m/z* values.

We assessed small mass differences by peak width-estimation approach (Hawkes et al., 2016). Based on the datasets prepared for our detailed analyses (set of seven samples), the FT-ICR MS detected 328 full triplets distributed over the mass range m/z 195 – 611; of these, the Orbitrap was able to detect 72 (mass range *m/z* 225 – 593) across the whole sample set. The NELHA sample provided best test conditions due to its richness in heteroatom-containing formulae. The FT-ICR MS did detect in total 547 single peaks belonging to the list of all triplet signals in the dataset (based on the [CHO]N_2O_2 analysis), of which 213 represented full triplets (n = 71) and covered a mass range from *m/z* 297 – 611. In the respective Orbitrap data, we found 380 of the considered signals in total, of which 36 represented full triplets (n = 12). In the H5-3a (anoxic aquifer) sample, the total number of FT-ICR MS triplet signals was lower (440 in total, 36 full triplets, *m/z* 269 – 457), and the Orbitrap also resolved lower numbers (of

those considered, 337 in total, 8 full, *m/z* 297 – 357), in good agreement with our calculated upper resolving limit. For the relatively heteroatom-poor IHSS sample, the FT-ICR MS only found two full triplets (441 signals in total), of which the Orbitrap resolved one (of those FT-ICR MS formulae considered, 402 signals were found, and one triplet at *m/z* 325).
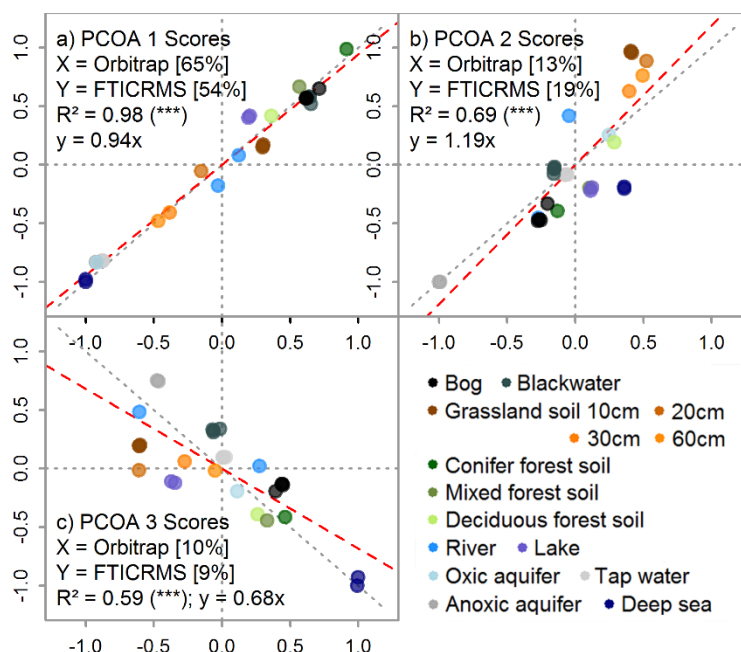
Figure 3-4. Multivariate agreement between FTMS instruments. Direct comparison of the first (a), second (b) and third (c) ordination scores for each sample determined by Principal Coordinates Analysis (PCoA). Note that axis notations are given within the figure, including the explained variability in percent. Replicate measurements are included (set of seven samples). Color indicates sample origin (see legend, in accordance with colors in Figure 3–1 and Figure 3–2).

3.2.4   Multivariate analysis of DOM molecular composition and trend retrieval

The separation of samples based on their molecular composition yielded highly similar results and the scores of the samples were highly correlated: The explained variance (R^2) of the scores decreased with coordinate rank from 0.98 over 0.69 to 0.59 (Figure 3-4). For the first coordinates, the regression slope did not differ significantly from the 1:1 line (within 95% confidence intervals). The first three axis explained 88 % and 82% of the summed variability in the datasets of the Orbitrap and the FT-ICR MS, respectively. The first coordinate of the Orbitrap data explained 11 % more of the internal variation (Figure 3-4a) and the second axis 6% less than the coordinates from the FT-ICR MS data (Figure 3-4b). IHSS, bog (N3B, N8B) and forest soil DOM types (H2S-5, T-5, W1-5, and JE-2-5-10) were clearly separated from deeper soil waters (JE-2-5; 20, 30 and 60 cm depth), aquifer (H5-3a, H3-2b), tap (TAP) or deep sea (NELHA) DOM.
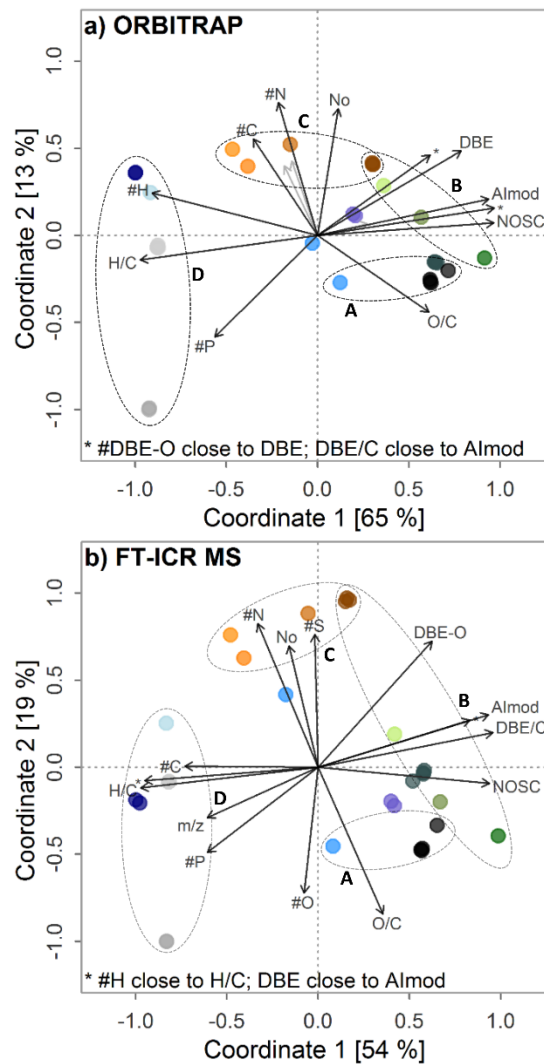
Figure 3-5. Post-gradient fitting on ordination obtained by a) Orbitrap and b) FT-ICR MS. Fitted gradients are abundance-weighted averages of indices, abbreviations: H/C, Hydrogen-to-Carbon ratio; O/C, Oxygen-to-Carbon ratio; AImod, Aromaticity index, DBE, double bond equivalents, DBE-O, DBE minus Oxygen atoms, DBE/C, DBE-to-Carbon ratio; NOSC, nominal oxidation state of Carbon, #C, #H, #O, #N, #S, #P, number of respective atoms. Circles A-D depicts sets of retrieved trends (see text) among samples. For color legend see Figure 3-4.

The derived linear gradients of indices (Figure 3-5) and molecular classes (Figure A-4) also were highly similar in the ordination plots for both instruments, confirming the results from the abundance patterns. In general, samples connected to positive scores on the first coordinate, like the bog, the black-water river IHSS and the forest soil water samples, were strongly linked to higher values of the aromaticity index (AI$_{MOD}$), the number of double bond equivalents (DBE) and the nominal oxidation state of carbon in the formulae (NOSC) and had larger amounts of formulae classified as black carbon-, polyphenol- and carbohydrate-like (BC, PP, SUG). Coincidently, high amount of hydrogen atoms and average H/C ratio were correlated negatively with the first coordinate. Subsequently, more formulae classified as "highly unsaturated" molecules, unsaturated aliphatics and peptide-like (HU, UA, PEP) were found in

aquifer, tap water and deep sea water samples. These groups of formulae show higher degree of saturation compared to aromatic or phenolic-like entities. The opposite trend of the chemical composition on the first coordinate was interestingly linked to different ion abundance distributions (Figure 3-1). Samples with positive scores on the first axis show a bi- or multi-modal distribution (Figure 3-1a – c) whereas samples with negative scores had a unimodal distribution with an average center at higher *m/z* values (Figure 3-1e – g). The separation on the second axis was mainly explained by the content of heteroatoms including oxygen and the number of formulae (Figure 3-5, Figure A-4). Besides general congruence in the observed separation pattern and linear trends explaining it both from the perspective of indices and molecular classes, we also observed clear differences in the ordination patterns. In the ICR system N and S had a positive loading and O and O/C a negative loading. In the Orbitrap system N also had a positive loading and O/C a negative loading, but O and S had no significant impact on the separation of the samples, which may explain the overall lower explained variance of the Orbitrap data on the second coordinate. Differences in multivariate response based on DOM composition were also linked to increased detection of mainly CHO formulae at higher m/z due to differences in fine-tuning of both instruments used in this study (Figure A-5, Figure A-6, Figure A-7).

3.3 Discussion

3.3.1 Comparison of mass abundance patterns between the Orbitrap and the ICR analyzer

The agreement of mass abundance patterns and abundance-weighted indices have been investigated by several studies (Table A-11) and are believed to be instrument-dependent, hindering direct inter-comparison (Pomerantz et al., 2011). The spectra described herein showed a good fit between both instruments (Figure 3-2, Table 3-2). Small deviations in congruence among mass abundance patterns of bog or IHSS samples agree with previous reports of similar effects for dystrophic lake water (Hawkes et al., 2016), but had no large effect on multivariate separation and derived conclusions (Figure 3-5).

Both instruments yielded highly similar responses, as can be seen from overall distribution of ion abundance (Table 3-2), with very similar amounts of information covered by assigned formulae in the mass range *m/z* 200 – 650 (70 – 80%), and a majority of ion abundance being commonly detected (>90%, up to 99% in single cases for the Orbitrap). The higher number of unassigned Orbitrap signals above *m/z* 200 can be largely explained by slight differences in

sensitivity of our Orbitrap method in the higher mass range, which may be due to differential tuning and instrument design, but may also point towards higher probability of taking false positives (Noise) into account (Table 3-2). However, only signals detected more than once in replicates were included, which strongly minimizes this type of error (Riedel and Dittmar, 2014). Our observations are thus subject to future and ongoing work within our group, to further improve the method (Figure A-1, Note A–1). Higher numbers of peaks may also arise from the difference in concentration (10 ppm, FT-ICR MS; 20 ppm, Orbitrap), which was due to the exchange of the ICR analyzer cell by a more sensitive type (Table 1-1). Similarly, increased contribution of simple CHO and CHNO formulae caused an increased annotation of Orbitrap-"specific" formulae (Figure A-1), mainly linked to slightly higher sensitivity of the Orbitrap in the upper mass range. However, total ion abundance of those signals was often less than 3% and reached 5% only in case of JE-2-5-10 and IHSS samples. These formulae were treated as "falsely"-categorized specific formulae as they were part of the FT-ICR MS dataset, but not found within similar samples and thus, categorized "specific". After subtraction of the "falsely" categorized specific formulae, the remaining ion abundance was close to being insignificant (Figure A-1) and thus was not further taken into account.

Together, our results indicate that Orbitrap deviations in ion abundance of common formulae (Figure 3-2) might be linked to sample-specific DOM constituents that influence ionization and detection in this FTMS analyzer. Samples that were less affected by fresh and decomposing organic matter inputs, such as the marine deep sea sample (NELHA), showed a remarkably good fit in mass abundance patterns (Figure 3-1e – g). The NELHA sample did not show signal suppression effects that were previously observed in another Orbitrap system (Hawkes et al., 2016). Besides sample effects, bi- or multimodal abundance distribution could also indicate unwanted source effects, such as adduct formation, or biases in the detector. All signals shown in Figure 3-1 are common formulae defined by their simultaneous detection and annotation by both instruments in the same samples (holding >70 – 80% total ion abundance in both instruments, Table 3-2). In turn, differences in spectrum shapes among instruments seen here can only be caused by abundance variation. Furthermore, adduct formulae including chloride, the most common adduct in negative-mode ESI, would be detected at m/z values not within the 0.5 ppm mass error criterion or only in formulae with unlikely combinations of heteroatoms N, S and P (compositions were checked with MIDAS Formula Calculator v.1.2.6, National High Magnetic Field Laboratory, Tallahassee, USA). Dimer formation as a source of ion abundance variation was effectively suppressed by use of source-induced dissociation (SID).

Tests with the IHSS sample showed that the use of SID alone did not explain the presence of a lower *m/z* hump centered at ~ *m/z* 225. Intensities and number of detected signals below *m/z* 200 were indeed increased by SID. However, this was not reflected by a simultaneous decrease in number or abundance of higher m/z ions (< *m/z* 800), altogether pointing towards better ionization of low molecular weight compounds in DOM. Potential analyzer effects cannot be ruled out and need further insight as they seem to be inherent to specific samples and Orbitrap-type instruments but were not within the scope of this paper. As noted above, slightly better sensitivity of the Orbitrap method in the higher mass range led to increased rates of formulae assigned only by the Orbitrap (> *m/z* 400, mainly simple CHO and CHNO compounds, Table 3-2, Figure A-1). Although this effect is likely caused by differences in instrument architecture similar to above reported deviations in intensities below *m/z* 400, the underrepresentation of higher mass ions in the FT-ICR MS spectra is not a consequence of instrument capabilities but rather a tradeoff in adjusting and fine-tuning the mass spectrometer response to individual user needs, as these formula compositions are easily resolvable by FT-ICR MS performance.

Our results indicate a strong preservation of the ion abundance information in both instruments, with only slight deviations discussed above, pointing towards representation of inherent sample characteristics by ESI-FTMS instruments. Note that the measurement settings of each instrument were the same for all samples (Table A-1). By establishing an Orbitrap method that minimized differences in ion abundance representation (and thus differences in ionization and detection), we were able to estimate the analytical window determined mainly by lower resolution and detection limits. However, a perfect overlap in instrument response may be difficult to achieve and would require similar tuning for each "type" of sample, which is costly and inflexible, if many different samples are compared. Specific signals only observed by the Orbitrap were in all cases close to the noise threshold (Figure A-1). This implies that FT-ICR MS-generated formula lists can be used to annotate the big majority of peaks detected in typical DOM datasets obtained by the Orbitrap. Of course, the Orbitrap has to be fine-tuned for this purpose. As demonstrated, this can be achieved by tuning the MS response (mainly ion abundance distribution) of one or two representative samples or available reference material that span the environmental/experimental gradient of interest. Likewise, the baseline formula list for Orbitrap peak annotation could be obtained based on a representative but small subset of samples.
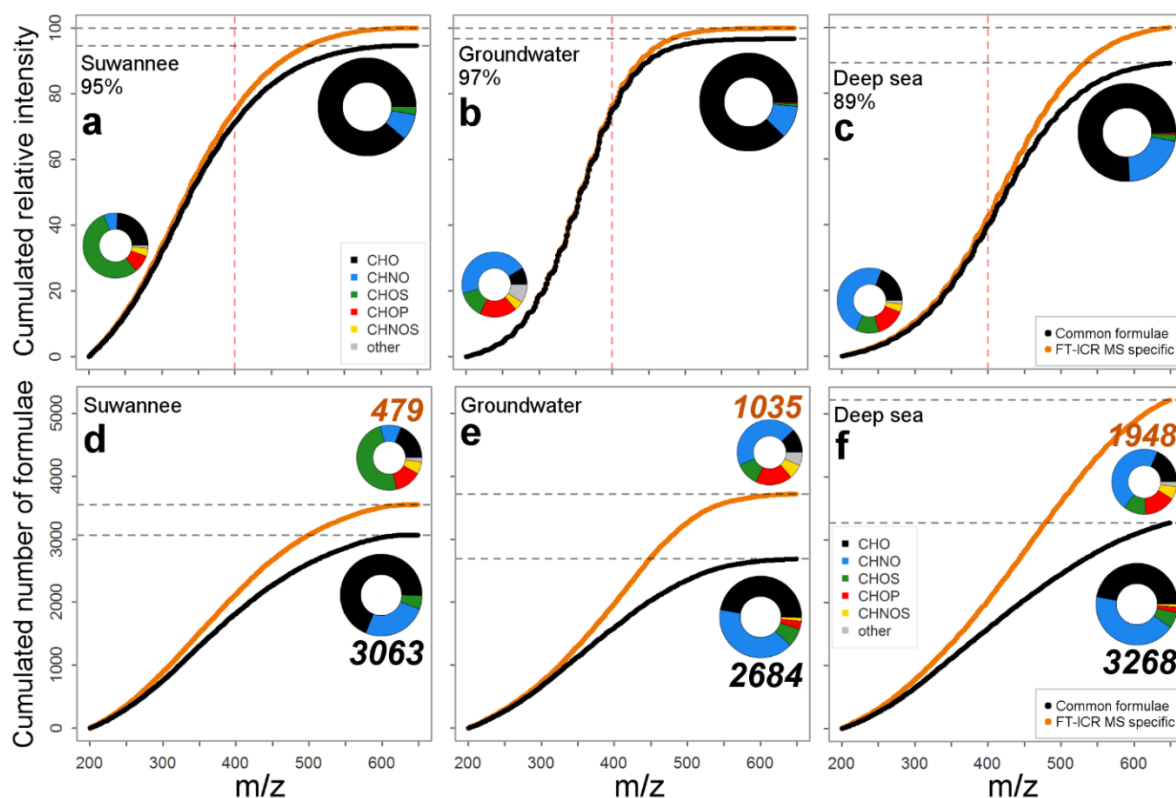
Figure 3-6. Comparison of MS coverage in terms of ion abundance (upper panels a – c) and formula number (lower panels d – f). Cumulated curves indicate a gradual increase of information gain (through FT-ICR MS in orange, vs. black line representing commonly detected formulae) in formula number and abundance information. The red vertical dotted line marks ~ m/z 400 where resolving power increasingly determines full coverage. Formula class ensembles covered by common and FT-ICR MS specific formulae are shown for the three specific samples in pie charts, indicating contribution to allover ion abundance (panels a – c) or formula number (d – f); the bigger pie always refers to the common set. Color encodes for compound class. CHNO includes N1-4, CHOS includes S1-2, "other": CHNOP/ CHOSP formulae.

3.3.2   Interplay of sample type and instrumental resolution on the gained information

In this study, we observed distinct clusters of formulae that were outside of the Orbitrap's analytical window as determined by FT-ICR MS (Figure 3-3). The degree to which these formulae groups contributed to the gained information depended strongly on sample type. This observation was clearly linked to the general shift in abundance distribution (Figure 3-1; Figure 3-3). A higher contribution of additional information by the fourth cluster translated into a stronger ion abundance deviation observed for the marine sample (11% of FT-ICR MS total abundance, Figure 3-6c). The region of the fourth cluster contributes substantially to a group of abundant signals that are ubiquitously found (Kellerman et al., 2014; Lechtenfeld et al., 2014; Roth et al., 2014) and are assigned to a maximum number of possible isomers (Hertkorn et al., 2007; Zark et al., 2017). Therefore the full complexity of marine samples is not fully captured by recent Orbitrap instruments. Strong similarity of ion abundance patterns suggests that reasons for information loss are mainly due to limitations in resolving power and not sensitivity (i.e., detection limit). This is especially true for compounds of the fourth cluster, as this region usually contributes to abundant signals. In contrast, the allover contribution of FT-ICR MS-specific formulae to total ion abundance (FT-ICR MS data) only accounted for five and three percent in the IHSS and H5-3a, reflecting additional detection of more peripheral compound groups (clusters 1 – 3, Figure 3-3a) or lower numbers of formulae associated with the fourth cluster (Figure 3-3c). Our results indicate that the better representation of DOM chemodiversity by FT-ICR MS is sample-specific and related to distinct groups of similarly composed compounds. Low-molecular-weight signals, indicative of fresh inputs of organic matter (Roth et al., accepted), were well captured by the Orbitrap. These results demonstrate the suitability of Orbitrap instruments for studying the individual composition of terrestrial DOM and its transformation during early stages of decomposition.

An assessment of the formulae lost from the Orbitrap's analytical window showed additional indication of sample-specific differences among instruments linked to DOM complexity. The group of commonly detected formulae was represented by a majority of CHO, CHNO and CHOS compounds (Figure 3-6d, e and f), and FT-ICR MS-specific formulae often belonged to additional heteroatom-containing compound classes (e.g., CHNOS). We observed a clear gradient of FT-ICR MS-added information from NSP-poor to increasingly NSP-rich samples (Figure 3-6a, b and c). Likewise, the total number of additionally detected compounds increased, together with their relative contribution to the total formula number: IHSS (13.5 %)

< H5-3a (27.8 %) < NELHA (37.3 %). This was also reflected by a parallel shift towards higher average m/z. However, the increased resolving power of modern Orbitrap instruments allowed detection of a wide array of CHNO and CHOS compounds, and even small numbers of CHOP and CHNOS formulae (Figure 3-6d, e and f insets, big pie charts). The Orbitrap thus can identify samples with more diverse heteroatom formula ensembles worth an FT-ICR MS investigation (Figure 3-6d, e and f insets, small pie charts).

However, the additional FT-ICR MS information made up only a small portion of the overall ion abundance (Figure 3-6a, b and c, max. 11%), confirming previous findings (Table A-11). The majority of geochemically important formulae were retrieved, but low-abundance peaks, possibly encoding to a higher degree for ecosystem-specific processes and nutrient dynamics (Kellerman et al., 2014; Roth et al., 2014), were captured more comprehensively by the FT-ICR MS. The correct abundance representation of common formulae (Figure 3-1, Figure 3-2) and the distribution of additionally resolved compound classes above ~ *m/z* 400 (in accordance with *m/z* shift observed in H/C vs. *m/z* space; Figure 3-3b, d and f) point to information loss mainly due to insufficient resolving power above that threshold. A manual search for FT-ICR MS-specific formulae in the Orbitrap data showed that the majority of peaks were either lost due to insufficient resolution of small peaks (rather flat shapes) or overlap by larger, neighboring signals, both becoming increasingly important at mass ranges above *m/z* 300. Below this value, the preferential loss of a small group of signals with mass defects above 0.17 Da indicated low sensitivity of our Orbitrap method for these signals.

Our analysis of triplet formation proved the increasing capacity of the most modern Orbitrap instruments to resolve bottleneck mass differences in complex DOM mass spectra rich in heteroatoms. Surprisingly, the last triplet in the NELHA sample was found at *m/z* 413, well above the calculated resolution limit for the triplets at ~ *m/z* 361, probably due to the relatively high intensities obtained in this specific sample. In principle, the lower coverage of triplet peaks in the Orbitrap data may also be due to insufficient sensitivity for certain signals. As noted above, 1) general accordance of ion abundance patterns of common formulae and 2) a broad limit in detection of heteroatom-rich compound groups above ~ *m/z* 400 points towards the expected decrease in Orbitrap performance due to insufficient resolving power. Moreover, the missing NELHA triplet signals in the Orbitrap data often remained unassigned due to peak interference as pointed out above. A group of interfering signals subsequently identified as the ^{13}C-isotopomers of CHNO formulae influenced [CHO]H$_4$S peak annotation. This interfering

series of signals is found approximately 1 mDa away (towards higher *m/z*) from the prospective sulfur signal and is thus even harder to resolve by the Orbitrap. The balance of both peak intensities influenced centroid positioning during peak picking, which subsequently affects peak alignment and formula assignment.

Taken together, ion abundance pattern and DOM complexity, which have been found to be important aspects of sample type in our study, influence the degree to which Orbitrap instruments keep up with FT-ICR MS performance. These results are important for Orbitrap use in future DOM studies and are even more promising as nominal resolving powers of 480.000 are achievable with a beta version of the Orbitrap's instrument software allowing for increased transient length (Zhurov et al., 2013). By analyzing two heteroatom-rich petroleum samples of even higher complexity, Zhurov and coworkers showed that increased resolving power was sufficient in the lower mass range up to ~ *m/z* 500, being somewhat higher than on the previous generation of Orbitrap cells (~ *m/z* 300, Smith et al., 2012). FT-ICR MS is still needed for a comprehensive analysis, even more so with FT-ICR MS improvements in place (Table 1-1), but is less necessary when heteroatom-poor DOM samples are analyzed.

3.3.3   Retrieval of discriminating information and biogeochemical trends

Multivariate statistics revealed a separation of ecosystem DOM types according to their proximity to recent inputs of fresh or freshly decomposing organic matter (Figure 3-5). Bog (N3B, N8B), IHSS and topsoil water samples (H2S-5, T-5, W1-5, and JE-2-5-10) showed strong contribution of aromatic and phenolic-type formulae in the lower mass range. During potential passage through deeper soil, aquifers and downstream aquatic systems, these signal become less dominant and NOSC, AI_{MOD} and DBE of DOM decrease simultaneously with a shift from bi- or multimodal ion abundance patterns to unimodal ones (Figure 3-1). Similar trends in loss of aromaticity have been described in soil and river settings using different methods (Creed et al., 2015; Klotzbücher et al., 2016). This implies that freshly decomposing materials can be traced by their contribution of aromatic- and phenolic-type formulae (Roth et al., accepted). In short, Roth et al. (2014) described several trends in a study including different sets of ecosystem samples. We found reproducible separation of a subset of these samples (water samples from forest soils, grassland soil, bogs, and rivers) on both instruments. This encompassed a) differentiation of surface and soil water samples, which were paralleled by changes in pH and vegetation type (clusters A and B in Figure 3-5a; Roth et al., 2014, 2013),

b) retrieval of a depth trend found for grassland soils (JE-2-5 samples) including an increase in molecular similarity to river and marine samples with depth mainly due to vanishing aromatic-type formulae in the lower mass range (cluster C; Roth et al., accepted), and c) discrimination of anoxic (H5-3a) and oxic (H3-2b) aquifer environments (cluster D, Nowak et al., 2017; Schwab et al., 2017). Unique formulae of forest sites showed a higher aromaticity compared to unique formulae found in less acidic grassland sites, supporting the finding that aromatic signals < m/z 300 might be linked to fresh and less decomposed organic matter inputs (Roth et al., accepted). Contrastingly, unique formulae found in grassland sites indicated stronger contribution of N and S formulae to DOM. Similarly, we found a significant link between the number of N atoms and the second coordinate. A paralleling trend regarding the sulfur content of formulae was only observed in the FT-ICR MS analysis. This may depict the surplus of information gained by higher resolution, where subtle differences in heteroatom contribution are still better constrained. The major separation showed high degree of robustness and was driven by commonly detected signals. All in all, the strong differentiation among samples of the set was corroborated by the retrieval of previously published trends with extended sets of similar samples.

3.4   Implications

The Orbitrap Elite has proven to be powerful for the challenging analysis of complex DOM fingerprints and the associated subtle differences in molecular composition that need to be resolved. Depending on the sample type, the additional information gained by the FT-ICR MS was either minor for samples that held mainly CHO formulae and had a lower average m/z, like the IHSS sample, or somewhat higher for samples that were NSP-rich and had higher average m/z, like the marine NELHA sample. The Orbitrap provided similar information compared to the FT-ICR MS, with the obvious aspect of lower resolving power being the main limitation in higher mass range. Accordingly, Orbitrap performance seems sufficient for the analysis of terrestrial samples that show abundance maxima in the lower mass range, and might soon even overcome this limitation through further increases in resolving power. Even under these circumstances, the Orbitrap is able to reproduce observed trends in molecular composition and allows separation of ecosystem types based on their DOM fingerprints. Drawbacks in resolving power can be tackled by increasing the specificity for analytes of interest during extraction (Li et al., 2016a; Tfaily et al., 2015), ionization (Hertkorn et al., 2008; Hockaday et al., 2009) or instrumental detection (Cao et al., 2016; Sleno, 2012). Although not comprehensive, broad

trends in heteroatom content were also captured by the Orbitrap. With regard to drawbacks in resolution of heteroatom-containing formulae, the Orbitrap may thus still be used as a tool to inform researchers about which samples would require regular or even improved FT-ICR MS performance. Our study provides an improved baseline for application of lower-resolution instruments. The open-source software used herein and published algorithms for molecular formula assignment and data analysis (Kew et al., 2017; Leefmann et al., 2019) will likely assist more researchers in contributing to DOM research in future. With analytical developments being available already now, FT-ICR MS and Orbitrap systems are of utmost importance to reveal full detail of the molecular composition, the origin and dynamics of DOM in both space and time. We finally recommend the use of internationally recognized reference materials that will help to decrease the instrument dependent tuning factors and simultaneously increase the comparability of the retrieved ion abundance patterns.

4 Whitesands and their link to blackwater DOM evolution in the Rio Negro basin

4.1 Introduction

Tropical rivers rank highest globally in terms of discharge and DOC export: The Amazon (incl. Tocantins), the Congo and the Orinoco deliver 25% of the global discharge and 18% of dissolved organic carbon (DOC) to the Atlantic Ocean (Raymond and Spencer, 2014). Due to within-river DOC transformation to CO_2, CH_4, or biomass, actual export of terrestrial organic carbon is probably even higher (Aufdenkampe et al., 2011; Ward et al., 2017). Riverine DOC export of tropical forests is thus a crucial component that interlinks global carbon and water cycles (Alvarez-Cobelas et al., 2012; Webb et al., 2018). Although found ubiquitously, DOC represents a complex mixture of thousands of individual molecules (dissolved organic matter, DOM; Minor et al., 2014) that exerts large compositional variation and thus holds much potential information on ecosystem imprints and processing (Raymond and Spencer, 2014). Ultrahigh resolution mass spectrometry (FTMS) by Orbitrap is especially suited to reveal molecular detail in early decomposition stages of terrestrial organic matter (chapter 3). Holistic understanding of the factors that govern the release and transformation of DOM in soils and within the aquatic continuum is needed to read the molecular signatures that emerge at different scales, from the interconnected pores to the watershed (Hutchins et al., 2017; Kellerman et al., 2015; Lynch et al., 2019; Peyton Smith et al., 2017). This knowledge is crucial to both understand the functioning of terrestrial environments and to keep them functional for future generations (Ward et al., 2017).

Sandy, deeply bleached soils classified as podzols are widespread in the Rio Negro basin, occurring in depressions on elevated plateaus and their slopes but also in valleys were they form large riparian areas (Do Nascimento et al., 2004; McClain et al., 1997; Montes et al., 2011; Remington et al., 2007; Zanchi et al., 2015). Do Nascimento et al. (2004) divide the occurrence of podzols into those occurring in high-elevation plateau terrain with the dominance of riparian systems (mid Amazon basin, e.g., north of Manaus) and in low-elevation plateau terrain with widespread podzol peneplains (north and upper Amazon basin, e.g., west and northwest of Manaus). However, whitesand areas with distinct *Campina* forests underlain by podzols are also found in elevated plateau positions (Demarchi et al., 2018; Zanchi et al., 2015). Knowledge about these diverse podzol ecosystems in terms of landscape position and

biogeochemical role remain scarce (Do Nascimento et al., 2004; Fritsch et al., 2011; Ishida et al., 2014; Lucas, 2001; Pereira et al., 2015). For example, despite the general acknowledgement that the occurrence of whitesand areas and blackwater river formation are linked (Ertel et al., 1986; Goulding et al., 1988; Leenheer, 1980; Remington et al., 2007; Sioli, 1954), we know surprisingly little about molecular imprints and DOM markers draining these widespread whitesand areas. McClain et al. (1997) estimated that while only 2 g of DOC m^{-2} yr^{-1} were exported by clay-dominated plateau (Oxisol) systems, sandy *Campina* sites exported about 40 g-DOC m^{-2} yr^{-1} and predicted that upland ecosystem markers would thus be better visible in higher order (downstream) rivers within the Rio Negro catchment. The dark color of blackwater bodies favors relatively low light penetration and depletion of oxygen in cooler deep waters, both preventing DOC degradation by photochemistry or the decomposer community (Goulding et al., 1988), and early reports indeed showed low oxidative capacities of blackwaters (Sioli, 1954). McClain et al. (1997) also estimated that overall exports of DOC by clearwater rivers linked to typical Oxisol/ kaolinite landscapes had to be compensated to a large degree by additional inputs from DOC-rich riparian zones and wetlands, as their groundwater inputs only accounted for ~30% of DOC export under baseflow (dry) conditions. Later studies corroborated these rough estimates (Zanchi et al., 2015). Remington et al. (2007) showed that the sorptive capacities of sandy soils were substantially lower that those of their neighbouring Oxisol and kaolinitic soils, thus explaining higher DOC fluxes from whitesand areas. Through comparative analysis of water-soluble organic matter (WSOM) from different soil horizons of an area affected by ongoing podsolization processes, Bardy et al. (2011) found that WSOM from deep horizons of well-developed podzols reflected best the chemical properties (specific UV-absorbance, pH and electrical conductivity) of similar WSOM isolates from groundwater draining the site, and also the receiving rivers. The authors concluded that Campina sites export specific DOC markers and hypothesized that these markers would allow distinguishing their imprint within downstream river networks. However, since then, the molecular differentiation of these remarkable landscape signals and their downstream fate has not been investigated further in detail. However, two recent studies consistently showed that oxidized and highly aromatic formulae > 300 *m/z* are indicative of the Rio Negro watershed (Gonsior et al., 2016; Simon et al., 2019). These findings are in line with reported markers in DOM from soils beneath temperate conifer forests and waters draining boreal bogs, all of which are usually characterized by acidic pH and high DOC levels, same as waters of the Rio Negro (Roth et al., 2015, 2014).

Whitesands and their link to blackwater DOM evolution in the Rio Negro basin

Recent understanding of DOC export from soils to groundwater and aquatic systems assumes constant exchange of DOC with coexisting soil organic carbon (SOC; Kaiser and Kalbitz, 2012; Leinemann et al., 2018). DOC production is sustained by topsoil "dominant source layers" rich in decomposing plant biomass (Ledesma et al., 2015; Lee et al., 2018; Malik and Gleixner, 2013). "Disconnection" of decomposers and molecules constituting soil organic carbon (SOC) governs their stabilization in bulk and deep soils, which show dilute SOC stocks with apparent old ^{14}C age and a dominant microbial imprint (Don et al., 2013; Gleixner, 2013; Kästner and Miltner, 2018; Liang et al., 2017; Sierra et al., 2018). In a previous study (chapter 3), DOM from systems dominated by constant recycling and low DOC levels (deeper soil layers, groundwater, marine systems) showed a marked mass shift towards higher molecular weights in Orbitrap data, as compared to DOM samples from sites dominated by inputs of decomposing biomass. Such mass shifts in DOM are commonly observed during fungal decomposition of DOM which goes along with decoloration, DOC loss, and resynthesis of novel compounds (Waggoner and Hatcher, 2017; Zavarzina et al., 2018) and are also described for soil depth profiles (Roth et al., accepted). Differences in soil texture, aggregation and hydraulic conductivity thus emerge as main determinants of SOM and DOM recycling because they control the number and intensity of recycling steps – much similar to the number of theoretical plates in chromatographic separations – along a molecule's passage from topsoil to groundwater (Kaiser and Kalbitz, 2012; Remington et al., 2007).

We hypothesized that the relatively large export of DOM from well-drained sandy systems in the Rio Negro basin (McClain et al., 1997; Remington et al., 2007) would allow to differentiate sources of Rio-Negro-specific markers (Bardy et al., 2011; McClain et al., 1997). In order to link podzol headwaters and the blackwater mainstem, we analyzed water samples from lysimeters, piezometers, streams and deep wells from two contrasting catchments in the elevated plateau region north of Manaus: A dry upland system characterized by smooth slopes and *Campina* forest cover, and a valley system typical for riparian, hydromorphic podzols associated with relatively steep slopes to neighboring plateau areas. We specially addressed the effects of ecohydrological differences (microclimate) and soil properties (texture) onto DOM composition. Finally, we compared the site-specific signals to reported sets of potential phenolic Rio Negro markers (Gonsior et al., 2016; Simon et al., 2019). Specifically, we adressed the following questions: 1) How do contrasting whitesand ecosystems differ in terms of their water chemistry (pH, electrical conductivity, water isotopy) and carbon biogeochemistry (DOC, carbon isotopy)? 2) How do clayey plateau and whitesand sites differ

in their molecular DOM composition, and how does it relate to water chemistry? 3) Do we find a link between ecosystem markers of whitesand sites and watershed-specific Rio Negro markers?
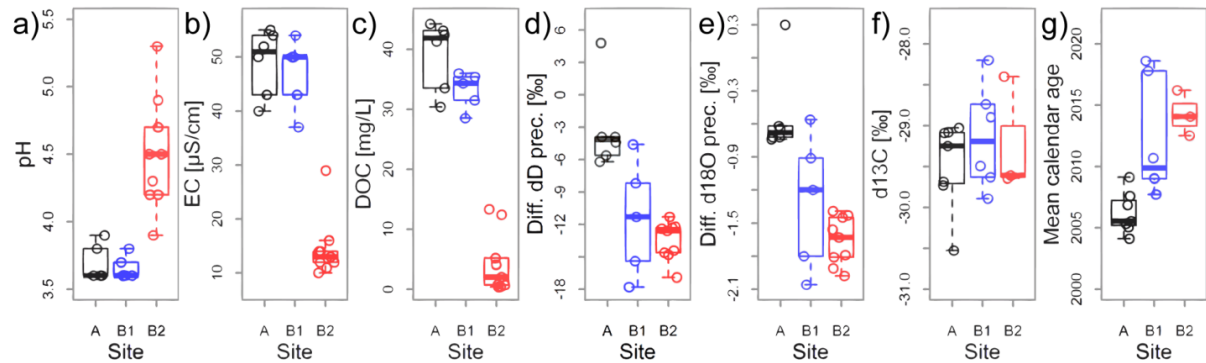
Figure 4-1. Overview of differences in water chemistry, water and carbon isotopes. Shown are the water chemistry (panels a-c), isotope chemistry of water samples (d, e) and their respective DOC extracts (f, g) from two contrasting whitesand areas (A, Reserva Campina; B, Reserva Cuieiras). Group B is separated into samples with high DOC (B1) and low DOC (B2). a) pH, b) electrical conductivity, c) dissolved organic carbon concentration, d) difference in δD values of water compared to regional average precipitation (OIPC estimate), e) difference in $\delta 18O$ values of water compared to regional average precipitation (OIPC estimate), f) delta13C of DOC extracts, g) mean calendar age based on F14C (Fraction Modern) values of DOC extracts. Note that groups size of B2 is smaller in f) and g) due to low C amounts of several samples (Tab. 1) and contains data from four samples measured in duplicates (E1, PR07, Rio Aşu, Rio Campina).

4.2   Results

4.2.1   Site characterization

The general analysis of water chemistry and carbon biogeochemistry showed clear differences among samples. The both sandy areas were generally characterized by similar water chemistry, but some samples at the Cuieiras site showed contrasting properties reflecting the presence of clayey Oxisols and Ultisols in the catchment (Table 4-1, Figure 4-1a – c). Acidity (pH), electrical conductivity (EC) and concentrations of dissolved organic carbon (DOC) were on average ~ 3.6, ~ 50 $\mu S*cm^{-1}$ and $34 - 42$ $mg-C*l^{-1}$ in blackwater samples from both sites (Figure 4-1). Samples influenced by lower blackwater influence at the Cuieiras site however showed low DOC levels ($\sim 1mg-C*l^{-1}$) with the exception of "intermediate samples" form piezometer PT06, lysimeters E3 and E4, and the stream (Rio Aşu), showing DOC levels in the range of 5 - 15 $mg-C*l^{-1}$ (Table 4-1, Figure 4-1). While the EC was consistently low in these samples ($\sim 11\mu S*cm^{-1}$), they indicated lower acidity: On average, pH was around 4.5 and highly variable (up to pH 5.3 in sample E3). The water isotopes indicated marked difference between sites (Table 4-1, Figure 4-1d and e, Figure 4-2). Samples from the Cuieiras site showed consistently lighter (more negative) δD and $\delta^{18}O$ values but all samples plotted on the local

meteoric water line (Figure 4-2) and were overall lighter as compared to local precipitation in previous years in the same months. While the δ^{13}C signatures of all DOC extracts (SPE-DOC) were roughly similar (Figure 4-1f), their calibrated radiocarbon ages were unexpectedly young – less than 2 years to 13 years at maximum – and differed markedly by 6 – 7 years between both sites (Figure 4-1g). SPE-DOC was younger in the Cuieiras site: youngest calendar ages were found in samples E1, PT06 and Rio Aşu. Highest calendar ages were found in left bank groundwater samples from Reserva Campina (P6, P7, Figure 2-2a).
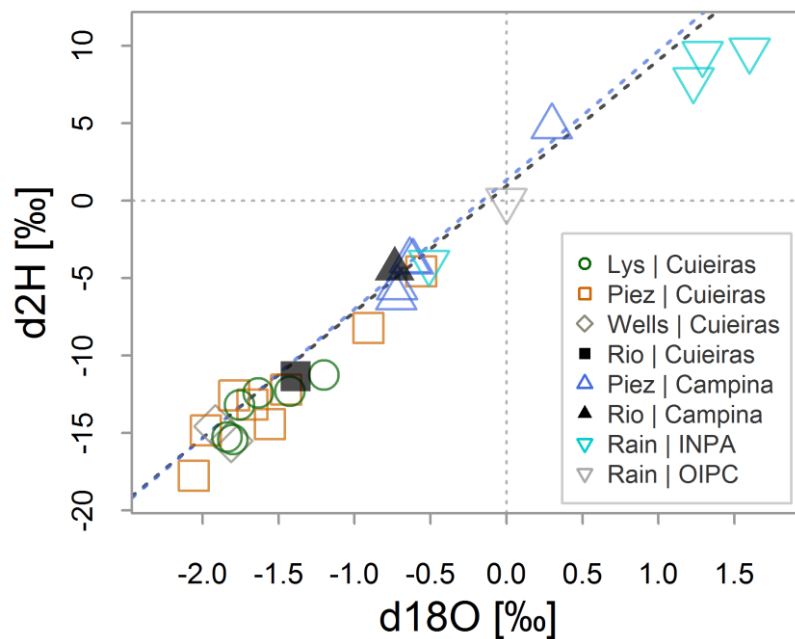
Figure 4-2. Water isotope data in relative notation against predicted average precipitation. Average precipitation taken from OIPC (OIPC 3.0 estimate, see methods for explanation) in all water samples taken in October/ November 2017. Black dotted line: Local meteoric water line (LMWL) constructed from data collected monthly by the INPA climatology station, located in the Adolpho Ducke Forest Reserve, Manaus, Brazil (data 08/15 - 02/17): d2H = 8.343*d18O + 13.362 (r=0.99, n=19). Four selected rain data points from October and November 2015 and 2016 are shown by cyan downward triangles. Grey dotted line: LMWL constructed from monthly dD and d18O in precipitation at IAEA/ WMO (International Atomic Energy Agency/ World Meteorological Organization) station in Manaus, Brazil, by Zhang et al. (2009), data from 1965 – 1990: d2H = 8.14 d18O + 12.96 (r=0.98, n=186).

Table 4-1. Combined data of samples from Reserva Cuieiras and Reserva Campina sampled in 2017. Greyed entries are denoting problematic data (see below table and text) related to contamination. EC, electrical conductivity; F14C Fraction Modern; $\Delta^{14}C$, correction accounting for decay between sample collection and measurement; SPE:DOC, ratio of mass of SPE resing and applied amount of DOC; EE, extraction efficiency based on DOC of samples and SPE extracts; Orbitrap, Quality of FTMS data.

ID	Date	Depth [1] [m]	pH	EC [μS/cm]	DOC [mg/L]	δ^2H [‰]	$\delta^{18}O$ [‰]	F14C	$\Delta^{14}C$ [‰]	Cal. Age [4] [yrs]	SPE:DOC	EE [%]	Orbitrap
PR11	11/01	0.96	4.2	14	0.4	-26.8	-5.03	n.d.[2]	-	-	2457	7	(ok)
PR10	10/31	n.d.	4.2	16	2.0	-24.2	-4.50	0.974	-33.6 [3]	2026 (1952)	994	73	(ok)
PR09	10/31	0.24	3.6	54	34.3	-20.2	-3.96	1.055	46.8	2009 (2010)	469	77	Ok
PR08	10/31	0.15	3.6	50	31.5	-16.6	-3.61	1.056	47.0	2009 (2010)	465	86	Ok
PR07	10/31	1.12	3.6	50	36.0	-29.8	-5.11	1.055	46.1	2009 (>2010)	456	76	Ok
PT06	10/31	n.d.	3.9	29	13.3	-26.4	-4.61	1.021	12.3	2016 (>2010)	500	78	Ok
PR06	11/01	1.86	4.5	11	0.4	-24.6	-4.84	n.d.[2]	-	-	4630	91	(ok)
PT09	11/01	n.d.	4.3	13	0.5	-25.2	-4.72	n.d.[2]	-	-	4404	12	Contam.
PP1	11/01	39.0	4.5	12	0.7	-26.6	-4.96	n.d.[2]	-	-	2190	33	(ok)
PP2	11/01	35.0	4.7	10	0.8	-28.9	-4.86	n.d.[2]	-	-	2443	15	(ok)
E1	10/31	1.0	3.7	43	35.4	-23.3	-4.25	1.011	2.7	2018 (>2010)	491	90	Ok
E2	10/31	1.2	3.8	37	28.5	-27.4	-4.85	0.984	-24.6 [3]	2024 (1955)	501	81	(ok)
E3	10/31	1.4	5.3	13	4.1	-24.4	-4.68	0.479	-524 [3]	-	1621	109	(ok)
E4	10/31	1.6	4.9	12	12.4	-24.3	-4.48	0.714	-292 [3]	-	596	87	(ok)
E5	10/31	1.8	4.7	11	0.9	-25.2	-4.81	0.054	-947 [3]	-	3703	200	Contam.
E6	10/31	2.0	4.5	12	1.0	-27.3	-4.89	n.d.[2]	-	-	3730	195	Contam.
Aşu	10/31	0	4.3	14	5.2	-23.3	-4.44	1.035	26.4	2013 (>2010)	487	95	Ok
P2	11/02	2.4	3.6	55	41.3	-7.2	-2.75	1.063	53.8	2008 (2008)	370	53	Ok
P4	11/02	1.5	3.9	40	30.4	-18.2	-3.77	1.055	46.4	2009 (2010)	332	67	Ok
P5	11/02	1.5	3.8	43	33.5	-15.9	-3.69	1.072	63.6	2006 (2005)	399	66	Ok
P6	11/02	1.5	3.6	50	42.4	-15.9	-3.67	1.079	70.5	2004 (2004)	362	68	Ok
P7	11/02	1.5	3.6	52	44.2	-17.6	-3.77	1.074	64.7	2005 (2005)	392	69	Ok
Rio	11/02	0	3.6	54	43.1	-16.4	-3.79	1.070	61.7	2006 (2006)	343	66	Ok

[1] R. Cuieiras: In piezometers, water level below surface at sampling (daily mean, hourly data), in lysimeters and wells: max. depth. R. Campina piezometers, max. depth. [2] n.d., not determined due to limited extract amount. [3] Value influenced by ^{14}C-dead contaminant signals. [4] Calendar age; first value obtained by prediction based on data from Graven (2015); second value in brackets obtained by OxCal 4.3 calibration with data up to 2010.

4.2.2   DOM characterization

The analysis of molecular DOM composition showed separation of previously described sample groups based on DOC concentrations, and revealed detailed differences among blackwater samples from both sites. Ultrahigh resolution mass spectra of DOM were evaluated by PCoA to extract the main coordinates representing maximum variability in the DOM signal (Figure 4-3). The PCoA separated DOM samples into two main clusters on the first coordinate (PCoA 1, Figure 4-3a, clusters 1 and 2; ~ related to groups B2 & A/B1 from Figure 4–1c) which held 72% of molecular variation and thus demonstrated the large difference of samples poor in DOC (cluster 1: PP1, PP2, PR6, PR11, PR10) from the remainder of the samples. However, there were notable deviations from the trend. DOM from relatively DOC-poor lysimeters E3 and E4 was also separated from these five samples (within cluster 2). The second coordinate separated the larger cluster 2 but the overall explained variability in molecular composition dropped sharply to 13%, indicating more subtle changes in DOM. Piezometers at the Campina site and lysimeters at the Cuieiras site (sub-cluster 2a) were separated from a sub-cluster (2b) of Ciuieras piezometer samples. In summary, the twofold separation of PCoA 1 and 2 allowed us to identify three possible DOM endmembers/ clusters: 1) DOM residing in DOC-poor groundwater environments; 2a) shallow groundwater with more constant saturated conditions, and 2b) DOM from soil with variable moisture (Figure 4-3a). The river samples, which represented the main flux of DOM export from both systems, reflected the imprint of these three clusters as well: Rio Aşu, the draining stream of the wetter Cuieiras site, plotted close to the main piezometer cluster 2a but was slightly shifted towards the direction of close by deeper groundwater samples (cluster 1, Figure 4-3a, Figure 2-2b); a result that was also reflected by relatively light water isotope values (Figure 4-2). The exported stream DOM however reflected better the composition of the shallow saturated zone of this specific valley system (piezometer cluster 2a). On the contrary, the draining stream of the Reserva Campina plotted in a position (Figure 4-3a, cluster 2b) in between its surrounding piezometer samples. The third coordinate spread cluster 2b samples by site (Figure 4-3b). In summary, our analysis revealed three subsurface DOM "endmembers" (cluster 1, 2a and 2b) at both sites.
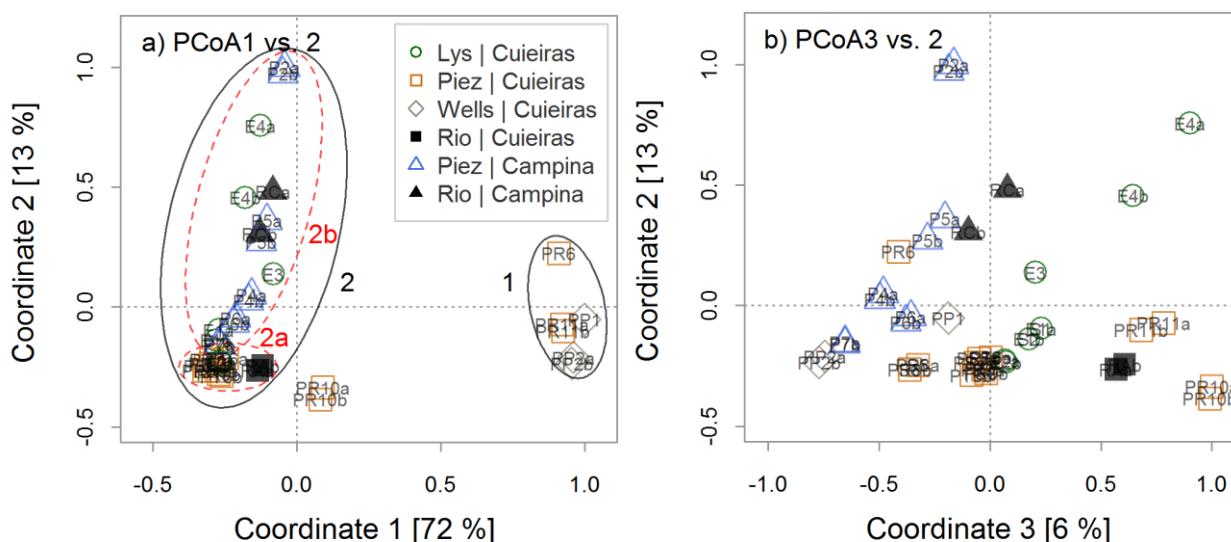
Figure 4-3. PCoA plots based on Bray Curtis dissimilarity of all available samples. The plots show separation of samples by a) first and second coordinate and b) third and second coordinate. The first coordinate marks the highest variability which separated plateau-influenced water samples (cluster 1) from rather valley-dominated samples in cluster 2 (containing also samples with mixed inputs, as found in most shallow piezometers). Coordinate 2 (PCoA 2) spread samples on a potential "moisture" gradient (see main text for explanation) into stable saturated (cluster 2a) and influence of changing moisture conditions (cluster 2b). Groundwater samples at Reserva Cuieiras reached a state (cluster 2a) that markedly differed from the surface signal influenced by wetting/ rewetting cycles (cluster 2b) but was largely different from groundwater of wells in deeply weathered clay-rich plateau soils (cluster 1). In summary, the analysis shows that the stream outflow DOM signal is dominated by an imprint stemming from surface processes in sandy areas, and that those signals seem to be influenced by moisture conditions. Cluster 2b was separated by site on the third coordinate (see appendix, Figure A-8).

The post-gradient ordination fitting analysis (*envfit* function, Figure 4-4) allowed us to relate main coordinates of separation to sample (i.e., pH) and DOM properties (i.e., aromaticity). The sample properties (Figure 4-1) reflected the main differences described above. Coordinate 1 (PCoA 1) was found to be linked to trends in DOC, EC and pH, whereas PCoA 2 was linked to differences in water isotopy. The main variability represented by the first coordinate was also found to parallel gradients of ion-abundance-weighted averages of molecular indices: The atomic H/C ratio was higher in DOC-poor samples (stronger saturation), whereas samples rich in DOC generally indicated the opposite, i.e., higher aromaticity, higher nominal oxidation state of carbon, higher carbon-normalized double bond equivalent, and higher atomic O/C ratio. A-priori-defined molecular groups (defined on the basis of the molecular indices) subsequently reflected the same trend: rather "labile" groups (D'Andrilli et al., 2015) showed higher dominance in samples of cluster 1 (i.e., PEP, unsaturated, O- and N-containing; HU, highly unsaturated; and UA, unsaturated aliphatics) whereas groups linked to oxygen-containing or aromatic structures showed higher dominance in samples of cluster 2 (i.e., PP, polyphenols; BC, black carbon-like; and SUG, carbohydrate-like).
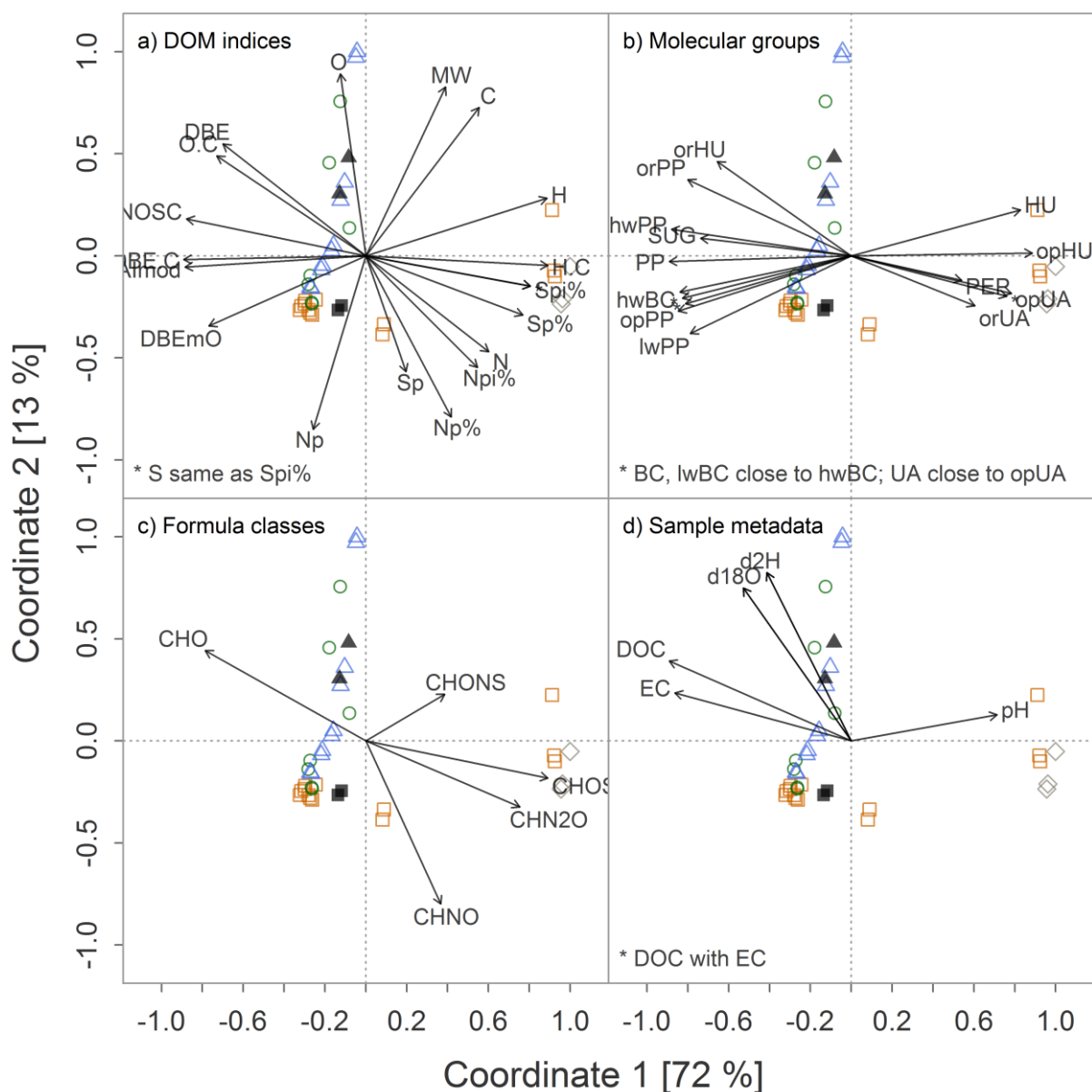
Figure 4-4. Factors governing PCoA separation of the whole set of samples. Same PCoA as presented in Figure 4-3a with added envfit analyses of different sets of variables: a) DOM indices, b) molecular groups, c) formula classes, d) sample metadata. See legend in Figure 4-3. Notes refer to variable arrow labels that were skipped for readability. Abbreviations: a) MW, molecular weight; Np/ Sp, number of N- or S-containing formulae; Np%/ Sp%, percentage of those based on total peaks; Npi%/ Spi%, percentage of those based on total ion abundance; C/ H/ O/ N/ S, number of atoms per formula; NOSC, nominal oxidation state of carbons; H.C and O.C, atomic ratios H/C and O/C; DBE, double bond equivalent; DBE.C, DBE based on C number; DBEmO, DBE minus oxygen number; AImod, aromaticity index. b) BC ("Black Carbon"), PP (polyphenols), HU (highly unsaturated); UA (unsaturated aliphatics); SUG (oxygen-rich, sugar-like); PEP (unsaturated, O- and N-containing); prefixes: lw ("low weight", formulae < 15 C atoms), hw ("high weight", formulae ≥ 15 C atoms), op ("O-poor", O/C ≤ 0.5), or ("O-rich", O/C > 0.5). c) CHO, formulae containing only C, H and O atoms; CHNO/ CHN2O, containing one or two N atoms; CHOS, containing one S atom; CHONS, containing one N and S atom. d) pH, acidity; EC, electrical conductivity; DOC, dissolved organic carbon; d2H, hydrogen isotopy; d18O, oxygen isotopy.

Figure 4-5. Factors governing PCoA separation of the reduced set of samples. Shown are the separation by coordinates 1 (71% explained variation) and 2 (15%) in a PCoA and subsequent envfit analyses of different sets of variables: a) DOM indices, b) molecular groups, c) formula classes, d) sample metadata. See legend in Figure 4-3. Notes refer to variable arrow labels that were skipped for readability. Abbreviations: a) MW, molecular weight; Np/ Sp, number of N- or S-containing formulae; Np%/ Sp%, percentage of those based on total peaks; Npi%/ Spi%, percentage of those based on total ion abundance; C/ H/ O/ N/ S, number of atoms per formula; NOSC, nominal oxidation state of carbons; H.C and O.C, atomic ratios H/C and O/C; DBE, double bond equivalent; DBE.C, DBE based on C number; DBEmO, DBE minus oxygen number; AImod, aromaticity index. b) BC ("Black Carbon"), PP (polyphenols), HU (highly unsaturated); UA (unsaturated aliphatics); SUG (oxygen-rich, sugar-like); PEP (unsaturated, O- and N-containing); prefixes: lw ("low weight", formulae < 15 C atoms), hw ("high weight", formulae ≥ 15 C atoms), op ("O-poor", O/C ≤ 0.5), or ("O-rich", O/C > 0.5). c) CHO, formulae containing only C, H and O atoms; CHNO/ CHN2O, containing one or two N atoms; CHOS, containing one S atom; CHONS, containing one N and S atom. d) pH, acidity; EC, electrical conductivity; DOC, dissolved organic carbon; d2H, hydrogen isotopy; d18O, oxygen isotopy; Cal.Age, calibrated calendar year (average DOC age); Site, binary code differentiating both sites.

Trends in the average number of C, O and N atoms per formula paralleled the separation of cluster 2a from the remainder of samples; consequently the percentage of CHO formulae was higher in this cluster whereas the percentage of CHNO formulae was reduced. Cluster 1 stood out in terms of sulfur-related formulae (percentage of CHOS peaks). Most tested variables were not linked to the third coordinate (Figure A-8); despite the obvious site-related variables DOC and EC, only the percentage of SUG formulae showed a significant correlation to the ordination plot and in the direction of the third coordinate, thus indicating higher values in Campina samples.

To test the relation of molecular properties of DOM and results of the radiocarbon analysis, we reduced our dataset for a repeated PCoA and *envfit* analysis (Figure 4-5). This was necessary as not all of our samples had enough C to be measured by AMS (Table 4-1). All of the samples included were part of the main cluster 2 in the first PCoA and *envfit* analysis (Figure 4-3, Figure 4-4) and thus our repeated analysis also allowed us to assess better the cluster-specific variability in DOM of cluster 2 (blackwater samples). The analysis revealed differences between sites mainly on the first coordinate (71% of explained variation) and showed a marked spread of samples on the second (15%), being related to a gradient from groundwater DOM to stream DOM which was consistent on both sites and also reflected partly the separation of sites by the third coordinate in our first PCoA (Figure 4-3b). Piezometer samples showed a greater variability in composition at Campina site, a result consistent with the previous analysis, and the composition of stream DOM reflected the remaining soil water samples in a similar manner as described above. The third coordinate showed only 5% of variation (Figure A-9) and was not further taken into account.

In order to test the relation between main coordinates and variation in properties of DOM (e.g., aromaticity) and samples (e.g., pH), we again performed a post-gradient ordination fitting analysis (*envfit* function, Figure 4-5). Separation of samples on the first coordinate was mainly linked to differences in average molecular weight (MW, based on *m/z* values), average number of O and C atoms per molecular formula, and average DBE and O/C ratio. Campina samples indicated higher values of all these indices; a trend observed also in the previous analysis. In contrast, Reserva Cuieiras samples showed higher number of peaks and were notably elevated in average N content per formula and thus percentage of CHNO and CHN_2O formulae. Along with trends in oxygenation, the oxygen-rich molecular groups of SUG and orPP showed highest dominance in samples from the Campina site, but also some intense CHOS compounds were especially elevated in these samples. Cuieiras samples were separated also by strong contribution of O-poor unsaturated aliphatics (opUA). The second coordinate was paralleled by a molecular trend in decreasing aromaticity

towards stream DOM (going along with significant trends of H/C ratio, NOSC and three of the molecular groups: HU, BC, PP). However, this trend was not as strong as the one observed in the previous PCoA of the whole sample set (Figure 4-3): Here, it was only related to the second coordinate (15% of variation vs. 72% in the first analysis). Interestingly, also the abundance and dominance of a second set of sulfur formulae was affected by this aromaticity gradient: The absolute and relative numbers of S-containing peaks were higher in piezometer samples most different from stream DOM; the class of CHONS formulae followed the same trend. The opposed trend in CHOS and CHONS formulae indicates differences in sulfur chemistry in CHO- vs. CHNO-dominated systems (Campina vs. Cuieiras, respectively). As expected for the separation of sites by the PCoA, *envfit* analysis of sample properties (DOC, EC) revealed the main differences reported beforehand (Figure 4-1) with no clear assignment of effects to one of the two coordinates. Solely the pH and the water isotopes showed a more pronounced correlation with the ordination (in the direction of the first and second coordinate, respectively). The ^{14}C information showed a similar trend and was thus inseparable of the site effect documented by other properties, as the pH or differences in CHO and CHNO formulae.

In an attempt to extract the formulae related the main clusters and blackwater samples from both sites, we conducted two Student's t-tests based on the relative ion abundances of formulae across samples of each group. In the first analysis, we compared cluster 1 (four samples, six mass spectra) and cluster 2 (16 samples, 31 mass spectra) from our first PCoA (Figure 4-3a) because they indicated the largest compositional difference (sample PR10 was included into the second cluster). In the following second analysis, we compared samples from cluster 2 that indicated separation of sites in the second PCoA/ *envfit* analysis (Figure 4-5) which was also correlated to apparent calendar age of DOM. This two-step approach allowed us to pin down the average molecular imprint of the three systems (clayey plateau samples, Campina blackwater, "riparian" Cuieiras blackwater) as revealed by sequential PCoA (Figure 4-6).
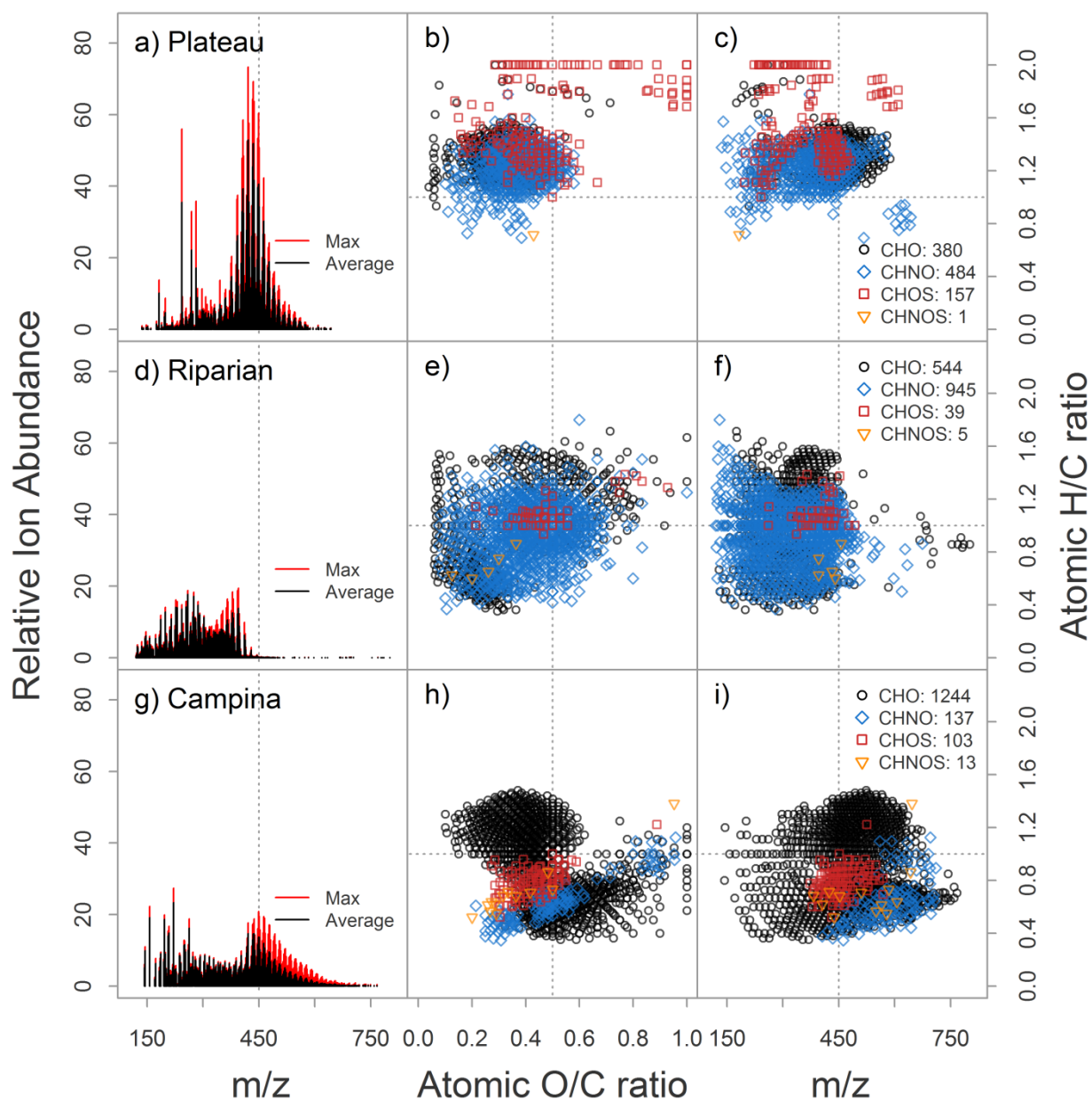
Figure 4-6. Indicative sets of marker formulae of three major landscape units in central Amazonia. Subsets of molecular formulae showing significantly higher ion abundance in groups of samples in cluster 1 samples (a – c; "Plateau", cluster1 defined in Figure 4-3a), in riparian samples of the Cuieiras site (d – f; "Riparian"), and in Campina samples (g – i). Sets of formulae specific to the whole cluster 2, formulae similarly abundant in clusters1 and 2, and those similarly abundant in riparian Cuieiras and Campina samples are presented in a similar way in Figure A-10. Panels a, d and f show the average and max mass spectra of each formula subset. Panels b, e and h show the formula subsets in Van Krevelen space (see formula classes and respective numbers of formulae in the legend in panels c, f and i). Panels c, f, and i show formula subsets in H/C vs. m/z space.

We found clear differences among clusters of samples. For example, samples of cluster 1 ("Plateau") that showed consistently low DOC and transparent color were differentiated by a set of formulae with a narrow mass and molecular formula distribution centered around *m/z* 425, O/C < 0.5 and H/C > 1 and showed a higher contribution of N-containing and S-containing formulae as compared to CHO-only formulae (Figure 4-6a – c) with a notable series of seemingly aliphatic sulfur formulae. In comparison, formulae indicative of cluster 2 (Figure A-10a – c) showed a much broader distribution up to *m/z* 800 and with a center at low *m/z* (~ 200 Da) and dominance of CHO formulae distributed across the whole Van Krevelen plot. A pool of distinct CHOS and CHNOS formulae centered at *m/z* 400 differentiated Cluster 2 from Cluster 1. Formulae above *m/z* 400 showed higher offsets in terms of average and maximum ion abundance, indicating larger differences within cluster 2 (Figure A-10a). The set of non-significant formulae present at similar ion abundance in both clusters from the first PCoA (Figure 4-3a) is presented in Figure A-10 (panels d – f).

To better differentiate samples of the larger cluster 2 (Campina blackwater/ "riparian" Cuieiras blackwater), formulae differentiating sites (as shown in Figure 4-5) were extracted in a similar analysis as described above. Clear differences between both sites were apparent through a sharp "cutoff" mass at *m/z* 400 (Figure 4-6d and g). Formulae indicative of samples from Reserva Cuieiras (named "riparian" due to the wet conditions at the site and for differentiation of the drier Campina site; Monteiro et al., 2014; Zanchi et al., 2015) were dominated by N-containing molecular formulae but generally showed a consistent distribution with previously described signals from cluster 2 besides the fact that many CHO formulae were non-indicative of this site. Part of them was highly indicative for the Campina site (Figure 4-6g – i): One cluster of CHO formulae was centered at O/C < 0.5 and H/C > 1, overlapping somewhat with the set of indicative formulae previously described for plateau DOM (Figure 4-6b vs. h), however with a marked absence of N-containing formulae and a higher average mass (Figure 4-6c vs. i). The second CHO cluster indicative of Campina samples was concentrated in the Van Krevelen plot area O/C > 0.3 and H/C < 1 and distributed across the whole mass scale (Figure 4-6h and i). Two smaller pools of CHNO and CHOS formulae were indicative too. However, a major number of formulae (~ 40%, n=1869) was not significantly different between sites and matched with the general envelop previously described for cluster 2 (Figure A-10a – c), among them also the distinct cluster of CHNOS formulae that thus seems to be a general feature at both sites. Together, these results showed a clear imprint of site-specific information in terms of DOM composition with subtle differences in contribution of clusters of CHO formulae and N- and S-containing formulae as seen in Van Krevelen and H/C vs. *m/z* space. Due to the strong correlation between sites and radiocarbon values, radiocarbon-correlated

molecular formulae showed similar patterns in Van Krevelen space (Figure A-11) as detailed above. A reanalysis conducted for each site separately (Figure A-12; Figure A-13) revealed that molecular formulae below *m/z* 400 showed a tendency to being present in seemingly older DOM whereas formulae above *m/z* 400 showed the opposite. This trend was consistently found for CHO, CHNO and CHOS formulae. At Reserva Cuieiras, formulae related to "older" DOM showed also a tendency to be less saturated (H/C < 0.8).

4.2.3 Whitesand ecosystems and their role in the formation of Rio Negro-specific DOM

Due to the surprising difference in DOM composition among the both whitesand areas, we compared their distinct DOM imprints ("Riparian" and "Campina", Figure 4-6) to published sets of Rio Negro-specific molecular formulae. With this, we aimed to gain a qualitative insight into the contribution of headwater stream DOM to the integrated signal of the Rio Negro watershed (Figure 4-7; Figure A-14). Despite the fact that many CHNO and CHOS formulae were part of the specific sets of signals in both whitesand sites, the realized overlap to specific Rio Negro DOM was restricted mainly to CHO formulae. Only in case of the broad analysis of data from Gonsior and coworkers, overlap with 238 CHNO formulae from riparian samples at Reserva Cuieiras was found. Both studies taken into account reported a notable contribution of aromatic compounds which thus formed a major part of the robust signal specific for Rio Negro samples. Our analysis revealed that the contribution of both headwater sites to these robust markers was close to exclusive: A distinct cluster of highly oxidized, aromatic formulae indicative of the Campina site dominated the Rio Negro-specific imprint independently reported by the two studies. Although the overlapping sets of formulae were found in the same area of the Van Krevelen plot (Figure 4-7a vs. c; H/C 0.4 – 0.8, O/C 0.4 – 0.8), this overlap was caused by different sets of molecular formulae in each study as indicated by the formula's *m/z* (Figure 4-7b vs. d), especially when only taking robust Rio Negro-specific signals into account. Gonsior and coworkers revealed a set of markers at higher mass (*m/z* 350 - 650), while our previous study revealed a set of markers in the range *m/z* 200 – 500. Nevertheless, all these molecular formulae were specific to the Campina site.
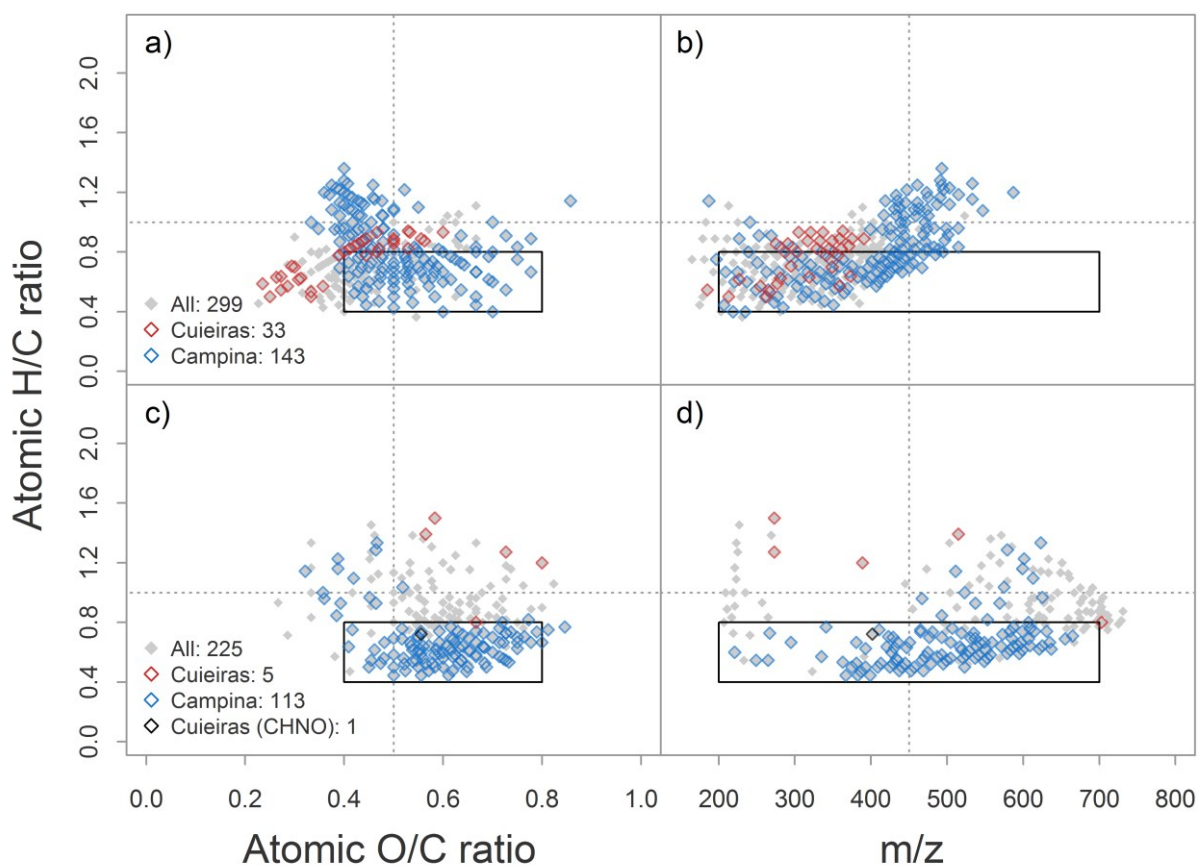
Figure 4-7. The molecular overlap between markers of whitesands and the Rio Negro. Overlap of site-specific formulae sets (see legend, CHO formulae in blue and red, CHNO formulae in black) with Rio Negro-specific formulae defined in a robust manner (see main text) as seen in Van Krevelen space (left panels) and H/C vs. m/z space (right panels). Small grey diamonds represent full sets of Rio Negro-specific formulae. Rio Negro-specific formulae were extracted from Simon et al., (2019) (a, b) and Gonsior et al., (2016) (c, d). The black box denotes the area of most robust specific formulae in Van Krevelen space (H/C 0.4 – 0.8; O/C 0.4 – 0.8) and is also transferred to H/C vs. m/z space (wide mass range of specific signals, m/z 200 – 700).

Data from blackwater samples of the two datasets (Gonsior et al., 2016, 18 measurements; Simon et al., 2019, six measurements) were merged with data from both whitesand areas to assess the degree of overlap between datasets and environments. Initial formula numbers were 5119 (this study), 4958 (Gonsior et al., 2016) and 3561 (Simon et al., 2019). Scan ranges differed slightly (*m/z* 120 – 1000, 180 – 800, 150 – 800), same as the range of detected signals (*m/z* 120 – 801, 180 – 799, 154 – 661), sample flow rates (7 µl/min, 2, 2), accumulation/ inlet times (100 ms, 200-500, 200), scan number (300, 500, 500) and presumably C concentration during electrospray ionization (ESI) in negative mode (20 mg/L in this study and in Simon et al. (2019), but not clearly stated in Gonsior et al. (2016)). Similar to the chosen ionization mode (ESI negative), resolution at *m/z* 400 was in the same order of magnitude (480k, 500k, 500k). The merged master list contained 7518 molecular formulae that represent an updated inventory of the Rio Negro watershed DOM spectrum. A Venn analysis of the whole dataset revealed a common set of 2091 formulae (, Figure 4-8a) and major

numbers of unique formulae in each dataset (common: ~28% of all formulae; specific to this study: 22%; spec. Gonsior et al. 2016, ~21%; spec. Simon et al. 2019, ~6%; Figure 4-8a). The similarity in CHO formulae was a little higher compared to the total set of formulae (35%; 16%; 21%; 3%; Figure 4-8b) whereas in terms of the CHNO formulae, similarity was lower than based on the total set (24%; 24%; 25%; 5%; Figure 4-8c). Clear differences among sample sets became evident in case of CHOS formulae, with no single formula being part of all three sets and most sulfur formulae being found in dataset C (Figure 4-8d). This differentiation in terms of CHO, CHNO and especially CHOS formulae was also revealed by cluster analysis based on presence and absence of formulae in individual samples (Figure 4-9). In general, the three datasets were clearly separated, and river samples were more close to another than to DOM from whitesand areas. Although covering large spatial gradients, the both river datasets were strongly uniform in their molecular composition as compared to the soil water samples which showed stronger variation at a much narrower spatial scale (cluster "a)" in Figure 4-9; Figure 2-1; Figure 2-2).
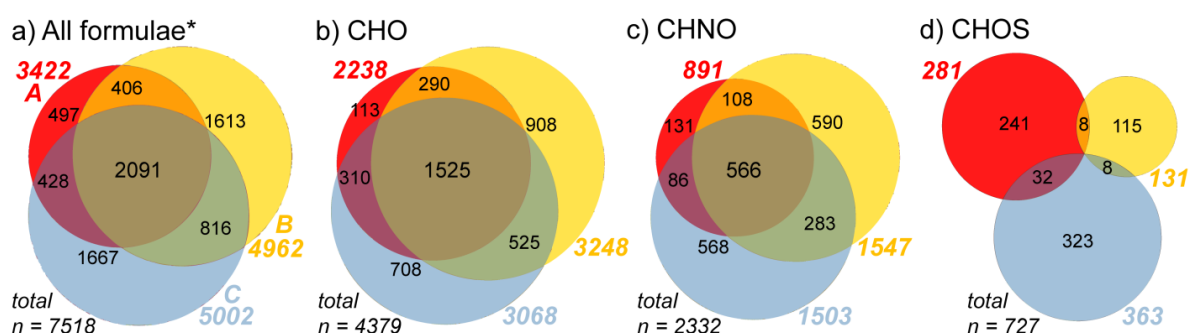
Figure 4-8. Comparison of molecular formula inventories of DOM in Rio Negro basin studies. Venn diagram showing the overlap in terms of molecular formulae in the three different FTMS datasets (A, red, Rio Negro and two tributaries, Simon et al. 2019; B, yellow, Rio Negro and lakes alongside the river , Gonsior et al. 2016; and C, blue, whitesand area dataset, this study).Panels show different sets of molecular formulae: a) whole set of molecular formulae; *asterisk: 80 CHONS formulae not included in panels c and d).b) Only formulae without Nitrogen or Sulfur atoms, c) Only formulae containing one or two N atoms, d) Only formulae containing a Sulfur atom.
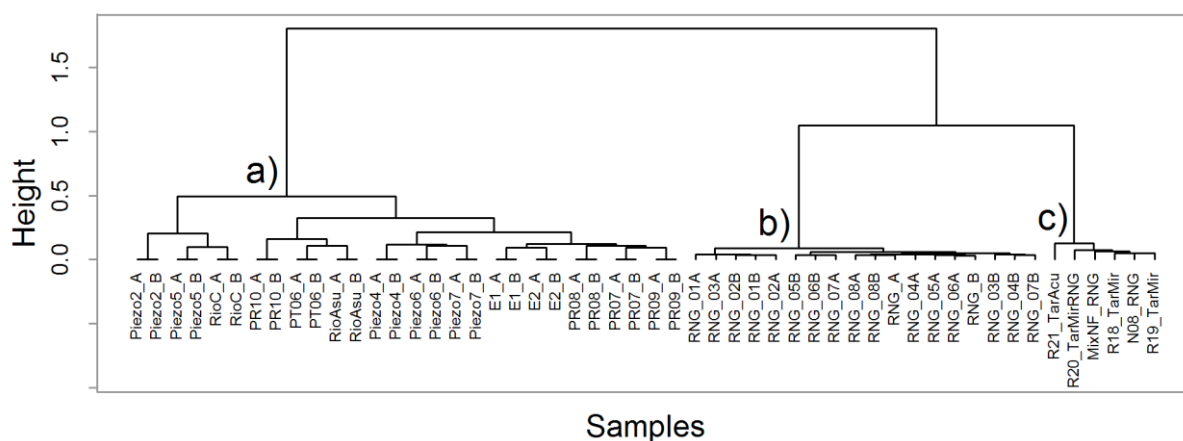
Figure 4-9. Similarity among mass spectra of Rio Negro basin studies. Result of the cluster analysis taking into account all formulae present in each measurement. Ion abundance information was omitted to reduce instrument-specific effects (such as tuning, ionization, etc.). Clusters: a) Samples from this study; b) data from Gonsior et al. (2016); c) data from Simon et al. (2019). Clustering was conducted in R Studio by function vegdist (with Bray-Curtis dissimilarity) of vegan package in R Studio and hclust (with Ward linkage, "ward.d2") of stats package.
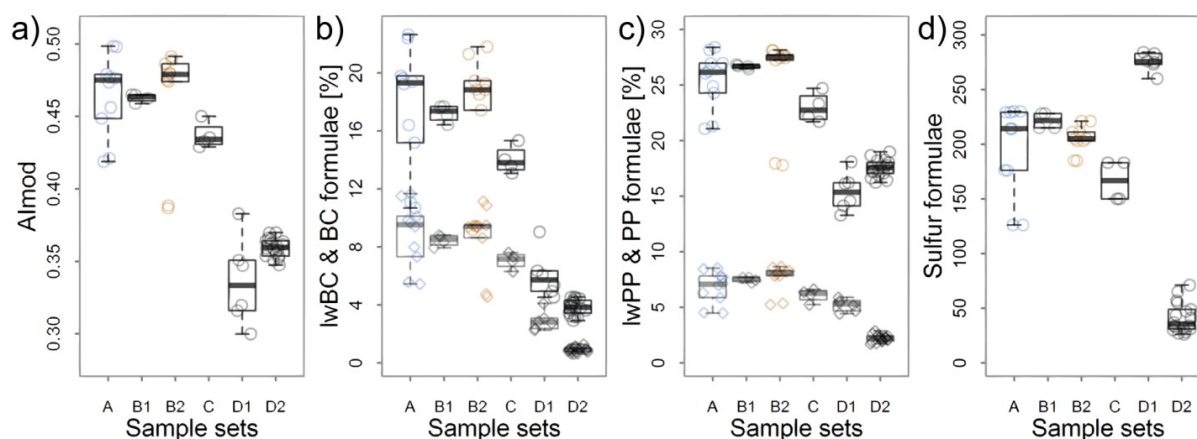
Figure 4-10. Evolution of DOM properties on a hypothetical hydrological gradient from soils to river. Our analysis took into account all available FTMS data of the Rio Negro basin: a) Aromaticity, b) relative numbers of polyphenols (total %, black boxes) and polyphenols with less than 15 C atoms (%lw + %hw = total, grey boxes), c) relative numbers of polycyclic condensed aromatics (Black Carbon-like, BC; total, black boxes) and BC formulae with less than 15 C atoms (%lw + %hw = total, grey boxes), and d) number of sulfur-containing formulae. Sample sets are ordered in the following way: (A) Campina piezometers, (B1) Cuieiras lysimeters, (B2) Cuieiras piezometers, (C) headwater streams (igarapés) of both whitesand sites, (D1) Rio Negro samples of the mainstem and smaller tributaries close to Manaus (Simon et al., 2019), and (D2) Rio Negro mainstem and adjacent lakes close to Novo Airão, 120 km NW of Manaus (Gonsior et al., 2016).

To test potential transformation of DOM between headwater sites and downstream higher-order rivers, abundance-weighted indices of the different samples were calculated and visualized according to hydrological order (i.e., from lysimeters to piezometers, streams, and rivers; Figure 4-10; Figure A-15). Despite differences in ionization and instrument tuning among studies, we found consistent decline in aromaticity and related parameters (H/C, DBE, not shown) at the transition from headwaters to higher order-rivers. Aromaticity declined from 0.45 to 0.35 and was reflected in a lower relative number of formulae classified as black carbon-like (BC) or

polyphenolic (PP), which declined roughly threefold (16% to 5% on average, Figure 4-10a) and twofold (27% to 17% on average, Figure 4-10c). Declines in low-weight black carbon-like formulae (strictly speaking, no polycyclic black carbon molecules due to less than 15 C atoms) were similarly high but less evident in low-weight polyphenols (Figure 4-10b and c, lower boxes) that showed a stable proportion as compared to our previous study (~5%) but showed strong decline in the data of Gonsior and coworkers (to ~2%). The contribution and number of sulfur formulae was highly variable within sample sets (Figure 4-6) and between sample sets (Figure 4-8d) and thus showed no clear gradient across the aquatic continuum (Figure 4-10d) but seemed to decline from soils to groundwater and streams, and showed a high degree of variability in higher order streams. Rio Negro-specific formulae showed the same trend (Figure A-15).

4.3 Discussion

4.3.1 Whitesand sites differ in their eco-hydrological characteristics

The study sites (Figure 2-1; Figure 2-2) showed remarkable differences in water chemistry that were mostly related to the presence of clay-rich soils at the Reserva Cuieiras (Figure 4-1). Despite the general similarity in samples from rather sandy soil environments producing waters rich in DOC (Figure 4-1a – c; boxplots A and B1 vs. B2), site differences persisted in water isotopy and radiocarbon content of SPE-DOC (Figure 4-1d – e, g; boxplots A vs. B1 and B2). Acidity (pH), electrical conductivity (EC), and dissolved organic carbon (DOC) concentrations of the studied samples were in line with previous reports from the Cuieiras site, reporting on gradients in these variables from plateau, slope and valley piezometers distributed in the Aşu catchment (Monteiro et al., 2014) and comparable landscape settings featuring laterite (Oxisol) soils, valleys and whitesand sites with typical Campina vegetation (Bardy et al., 2011; Cerri et al., 1991; Do Nascimento et al., 2008). Moreover these studies also showed the strong relation between pH, EC and DOC: We observed lowest pH in sandy soil environments and a strong control of electrical conductivity by dissolved organic carbon in all settings (Do Nascimento et al., 2008; Monteiro et al., 2014; Waterloo et al., 2006). Plateau samples showed higher pH values and strong variability, probably indicating ongoing buffering by oxide dissolution (Do Nascimento et al., 2008, 2004; Markewitz et al., 2001). DOC concentrations of Rio Aşu indicated pronounced delayed flow (base flow) conditions at the time point of sampling at the Cuieiras site (Waterloo et al., 2006) while DOC concentration of the stream ("Rio Campina") draining the Campina site revealed similarly high levels as the surrounding piezometers which we interpret as a sign of the strong connectivity between the small headwater

stream and its surrounding soil and in line with previous observations (McClain et al., 1997; Zanchi et al., 2015).

The water isotope data indicated strong differences between sites and generally low enrichment in heavy isotopes (lighter) composition as compared to simulated average isotope composition of local precipitation. Available data points from measurements of precipitation at INPA's meteorological station at Reserva Ducke, Manaus, indicated even heavier isotope composition in 2015 and 2016 (Figure 4-2; Figure A-16). Water is isotopically light during the wet season (February – May), and becomes heavy in the dry season, peaking from July – September (Zhang et al., 2009) which propagates to water discharged by rivers (Leopoldo et al., 1982). River isotopic composition reflects mainly groundwater that is recharged in the wet season when rain is isotopically light (Miguez-Macho and Fan, 2012; Tomasella et al., 2007; Zanchi et al., 2015). Large amounts of water are thus stored in the soils and slowly supplied from adjacent groundwater, and losses of water due to evapotranspiration from soil are minimal (Kunert et al., 2017; Leopoldo et al., 1982). It is not surprising to find a relatively heavy isotopic composition of groundwater and stream water in the dry and more elevated Reserva Campina; more direct links between precipitation and discharge are expectedly higher due to the elevated position (Miguez-Macho and Fan, 2012; Zanchi et al., 2015, 2014). For example, Zanchi et al. (2015) suggests that the Campina site experiences more frequent drought events. Piezometer sample P2, located at the uppermost point of the transect (Figure 2-2a), thus shows a heavier water isotopic composition. It is concluded that this sample probably reflects most the isotopic composition of local precipitation at the time point of our sampling. Also samples PR08 and PR09 from Reserva Cuieiras are more enriched as compared to other samples at the site. Water level data from these piezometers indicated exceptionally high levels (~ less than 0.2 m below soil surface) probably related to percolating rain that shifts the isotopic composition of these samples to more enriched values. The shallowest lysimeter E1 and the river sample (Rio Aşu) indicated a similar influence.

Several samples were not measured for carbon isotopes due to insufficient amounts of isolated C (Table 4-1; n=6; PR10, PR06, PT09, PP1, PP2, E6) and some others showed high levels of contamination by ^{14}C-dead organic material (n=5; PR10, E2, E3, E4, E5), especially lysimeter systems ("E") which seem to be related to the used materials. This was also apparent in the Orbitrap mass spectra. In contrast to bulk ^{14}C analysis, the simple exclusion of contaminant peaks in the Orbitrap data still permitted subsequent analyses of the full sample set (see methods section). Carbon isotope data revealed contrasting properties of DOC in the studied forest systems: While

δ^{13}C values were similar in all samples taken into account (Figure 4-1f) and reflected a uniform C3 source of organic carbon in line with other reports from the tropics (Häggi et al., 2016; Moyer et al., 2013; Sun et al., 2017), radiocarbon data, although generally indicating young DOC (< 13 yrs in cal. age), showed also clear differences among sites with ~ 6 – 7 years older DOC at the Campina site (Figure 4-1g). Young radiocarbon ages of DOC in the tropics have been reported frequently (Mayorga et al., 2005; Moyer et al., 2013; Ward et al., 2013) and have also been reported from soils (Doupoux et al., 2017; James et al., 2019): Although radiocarbon age shows the usual decline with depth, James and coworkers also revealed that water-soluble organic matter (WSOM) had a consistently younger age than the coexisting soil organic matter. This age offset increased with depth from 600 years at 5 cm depth and leveled off to rather constant ~ 1800 to 2500 years below 30 cm (up to 140 cm depth, in two native Cerrado stands). However, their extractions were conducted with alkaline solution (0.5 M K_2SO_4) that releases presumably older fractions of SOM (McClain et al., 1997). Although our results support this, the extreme young ages of purely water-soluble DOC at our sampling sites are a highly unexpected finding that may relate to strong linkages between C fixation, DOC release, and tight cycling of nutrients (Mayorga et al., 2005; Townsend et al., 2011). The observed slight but consistent differences in DOC radiocarbon ages between the drier and the wetter system likely relate to processes that govern the short-term (years to decades) turnover of organic matter (Zanchi et al., 2015). For example, Zanchi et al. (2011) showed that litter turnover was reduced in the Campina site as opposed to the riparian system; and Zanchi et al. (2014) show that the riparian system shows highest efflux rates of CO_2 while the Campina site shows lowest values. Lower DOC export due to missing adjacent groundwater inputs that dominate the flow of water through riparian systems could produce a more dominant imprint from topsoil layers (Raeke et al., 2017; Sanderman et al., 2009). Alternatively, aged DOC could also be released from former oxide minerals that are prone to dissolution under podsolization (Bardy et al., 2011; Do Nascimento et al., 2008).

The notable differences in extraction efficiencies among sites was probably not due to slight underestimation of DOC in Campina samples; although effects on DOM composition have been reported (Li et al., 2016b), we take the strong agreement between previously reported C/N ratios of similar DOC isolates from similar sites in the region (in case of the Reserva Campina, the same site) and the dominance of CHO and CHNO formulae in the different sites of this study as a sign of no large bias (McClain et al., 1997). Contamination however was an issue. We discarded ^{14}C and Orbitrap data of samples PT09, E5 and E6 (which were located in close proximity to one another). These samples showed very low DOC concentrations and were thus prone to contamination by

sampling equipment. Other samples were affected only in terms of radiocarbon data; their Orbitrap data was used after removal of highly abundant peaks that surmounted the usual Gaussian envelop observed in "clean" DOM mass spectra (samples marked with "(ok)" in Table 4-1). Sample low in DOC were especially prone to contamination; however, also samples from lysimeter systems ("E") indicated elevated levels of contamination through the sampling equipment. These samples were only included into our first broad analysis and interpreted carefully; for the pure blackwater analysis, we excluded all ambiguous samples and used only those samples indicating no contamination both in Orbitrap data and radiocarbon content (Table 4–1).

4.3.2 Blackwater DOM reflects differences in eco-hydrology

For DOM interpretations, DOC and water isotopy were used as proxies of soil texture (Remington et al., 2007) and ecohydrology (Kunert et al., 2017; Leopoldo et al., 1982; Tomasella et al., 2007). DOM reflected the main differences that were also apparent from water chemistry and water isotopes (Figure 4-3, Figure 4-4, Figure 4-5). The most striking difference in molecular composition between samples was related to texture (clayey plateau soil vs. sandy valley and Campina soils; cluster 1 vs. 2). A-priori defined molecular groups paralleled this separation (Figure 4-4b). The major differentiation in terms of a mid-molecular weight cluster rich in nitrogen and hydrogen, but poor in oxygen, reflected these molecular group assignments (HU, UA, PEP; Figure 4-6a – c), while molecular formulae dominant in cluster 2 samples were less saturated and mainly of the CHO class (Figure A-9a – c). Clayey materials are often associated to N-containing compounds (Aufdenkampe et al., 2001; Chassé et al., 2015; Lutfalla et al., 2019; Newcomb et al., 2017). Differences in sorption capacities and water retention of soils found in the Rio Negro watershed have been described by many authors (Marques et al., 2004; Remington et al., 2007), and showed that lower C content in subsoils goes along with reduced flow, i.e., stronger water retention (Marques et al., 2010). Bailey et al. (2017) found strong differences in molecular composition of SOM and DOM from differing pore domains; this suggests that soils differing in pore size distribution, and thus water retention, would also reflect these (Bailey et al., 2017; Vogel et al., 2014). It is intriguing that Bailey et al. (2017) found the exact opposite of the trend described above, namely that aromatic compounds (lignins, tannins) were preferentially drained when stronger suction was applied, thus draining smaller pores (Bailey et al., 2017) which might be due to the rather sandy texture of their chosen soils. In our study, N-containing and aliphatic compounds were indicative of clay-dominated sampling sites, and suggested that low DOC levels go along with strongly recycled, mainly microbial DOM (Kaiser and Kalbitz, 2012; Kästner and Miltner, 2018; Roth et al., accepted) which

is also in line with results presented in chapter 3 (recycled DOM is low in plant-related phenolics). We conclude that the overall specificity of rather saturated N- and S-containing compounds in plateau soil environments probably reflects the remains of a decomposition spiral that turns on ever decreasing rates with distance from inputs (top soils, lateral flow), in line with recent models and observations of DOM cycling along depth (Bardy et al., 2011; Kaiser and Kalbitz, 2012; Leinemann et al., 2018; Roth et al., accepted) and SOM preservation through "disconnection", i.e., dilution (Don et al., 2013; Gleixner, 2013; Sierra et al., 2018). As a result, the remains of the decomposer community become ever more dominant in the DOM pool because SOM is catabolized at slow rates but constantly exchanges with the DOM pool as suggested by Kaiser & Kalbitz (2012). This process would be even more accelerated (i.e., evident at shallower depths) if contact times and surface area would be maximal as in clayey soils, and also reflect the reported slower growth rates of microbes and turnover of carbon in low-substrate environments/ small pore domains (Don et al., 2013; Heitkötter and Marschner, 2018; Juyal et al., 2018). Oxygen limitation in such small pores could stop the turnover at an even earlier time points (Borer et al., 2018), as has been shown in aquatic sediments (Boye et al., 2017). The prevalence of N- and S-containing molecular formulae showing a relatively high degree of saturation ($H/C > 1$) and a low degree of oxygenation ($O/C < 0.6$) could thus also be a sign of anoxia which has been reported in blackwater rivers (Sioli, 1954).

DOM also reflected differences that were linked to the frequency of drying/ rewetting events (non-saturated/ saturated conditions). Samples from the drier Campina site and surface-influenced samples from the Cuieiras site (lysimeters; cluster 2b) were separated from cluster 2a which represented a common DOM signature of riparian (sand-dominated) piezometer samples (Figure 4-3a). Here, water isotopes were used as a proxy to relate DOM properties to the ecohydrological setting. Differences described for water isotopes were thus mostly reflected by the second coordinate of the PCoA, but also superimposed by a site effect (similar to radiocarbon variation). Largest differences between both clusters were lower weighted means of molecular weight and contents of oxygen and carbon, and higher numbers of N- and S-containing peaks in the riparian samples from Reserva Cuieiras (Figure 4-4), a results showing high consistency with previous reports on C/N ratios of DOM, partly from the same site (McClain et al., 1997). We assume that differences in water levels or water level fluctuations would affect the release and oxidation of organic matter (Peyton Smith et al., 2017; Schimel, 2018), as opposed to the above described reduced N- and S-containing formulae indicative of deep plateau soils. It is especially intriguing that McClain and coworkers find an increase in C/N ratios from upland soil (10) to riparian (15) and stream DOM (20), in line with the strong differences in CHNO fingerprints between these

landscape elements in the Cuieiras site (Figure 4-6a – f). Several authors have shown that especially upslope positions in elevation gradients in the region experience strongest fluctuations in water levels with deepest water levels at the end of the dry season (Do Nascimento et al., 2008; Lucas et al., 2012; McClain et al., 1997; Monteiro et al., 2014; Patel-Sorrentino et al., 2007; Saunders et al., 2006; Tomasella et al., 2007; Zanchi et al., 2015). Unpublished data from S. Ferreira, INPA (available at https://www.researchgate.net/ publication/242223718) shows that water contents differed between landscape elements of Reserva Ducke north of Manaus, which shows similar features to the environments studied here: Slope soils showed highest and most stable moisture content during dry and wet phases of the year with ~50 Vol-% as compared to plateau and valley soils. Clayey plateau soils showed similarly stable but slightly lower moisture content (30-40 Vol-%) and sandy valley soils were characterized by stronger variability between wet and dry phases (5-35 Vol-% and steepest depth gradient). These observations support our finding of DOM showing an imprint of dryness; especially as we sampled at the end of the dry season. Such seasonal differences in DOM composition and/ or contribution of different headwaters could probably also affect differences in whole-watershed DOM signatures, as reported in a previous study (Simon et al., 2019), with apparently higher levels of aromaticity after peak discharge of Rio Solimões and Rio Negro in July 2014 as opposed to October 2013. This remains to be tested by larger-scale studies.

We analyzed site-specific differences between blackwater samples in more detail in combination with radiocarbon data which can serve as a proxy of DOC "exchange" with recently-fixed CO_2 (Figure 4-5, Figure 4-6d – i, Figure A-9g - i). The PCoA reproduced the allover separation of clusters 2a and 2b (Figure 4-5a, c). However, the *envfit* analysis also revealed more subtle differences between sites, with groups of highly oxidized structures (SUG, orPP) and oxygen-poor unsaturated aliphatics (opUA) reflecting site differences in age, water isotopes, DOC, EC and pH (Figure 4-5b, d). The analysis of differentiating molecular formulae agreed well with the PCoA result (Figure 4-6d – i): Distinct higher-molecular weight clusters of CHO formulae (on average, $m/z > 400$), one of them a highly oxidized cluster (O/C > 0.4) with H/C levels below 1, formed the distinct molecular fingerprint of the dry Campina site. Riparian valley samples from the Cuieiras site showed distinct contribution of lower-molecular weight CHNO and CHO compounds ($m/z < 450$), in line with a strong contribution of CHNO compounds with similar size distribution but even higher saturation in plateau-derived DOM samples from the Cuieiras site (Figure 4-6a – c), as discussed above. It can be expected that riparian DOM fingerprints would probably reflect in part the lateral flows from the plateau; however, permanent saturation might contribute to the preservation of a wide suite of

organic compounds (Boye et al., 2017). Bardy et al. (2008) also report on the presence of phytotoxic compounds in waterlogged podzols which could also explain the preservation of nitrogen-containing dissolved organic molecules. Our analysis documented the imprint of riparian soil DOM in the signature of exported organic matter (stream DOM) and the simultaneous overprint of plateau-derived DOM signatures by this signal, despite a major contribution of water fluxes from replenished groundwater resources of the surrounding plateau landscape. Interestingly, it is this water flux that drives continuous export of young dissolved organic matter from the riparian valley systems at the Cuieiras site (Ledesma et al., 2015; Lee et al., 2018). The vast extent of Oxisols in the Amazon, their role in carbon turnover also at deeper horizons (as discussed above), and the dominance of groundwater in controlling discharge (Gross and Harrison, 2019; James et al., 2019; Miguez-Macho and Fan, 2012) together with our findings all indicate a potentially strong molecular imprint that should be distinguishable in the downstream river network. Given the presumably old, microbial signature of "persistent" DOM from plateau soils, one may guess that the amounts of aged (and possibly persistent) DOM that enter aquatic systems are underestimated; however this quantity will be overprinted by large amounts of young DOC exported from riparian systems (Reserva Cuieiras) and topsoil layers (Reserva Campina). The DOC ages observed in our study were very young and thus probably reflect the masking of a small pool of old carbon by a larger pool of recently-fixed C (Dean et al., 2019; Follett et al., 2014). Evidence for this hypothesis comes from the common observation of evasion of largely radiocarbon-young CO_2 in the downstream river network (Hutchins et al., 2017; Mayorga et al., 2005; Ward et al., 2013) but also from the observation that aquatic DOC becomes older upon respiration (Dean et al., 2019) along its passage in the river continuum (Kellerman et al., 2018), or during coagulation in water treatment (Raeke et al., 2017). All of these processes are in some way preferentially attacking "terrestrial" DOM, i.e., phenolic and aromatic molecules, and thus reflect results presented in chapter 3 (decrease in aromaticity during ecosystem passage). Our results support the view that the majority of exported whitesand terrestrial carbon is characterized by an age distribution shift towards modern values. However, the results also imply that the radiocarbon-old residual DOC found in higher-order rivers or marine systems after evasion of young C (Kellerman et al., 2018; Mayorga et al., 2005) may originate from deep plateau soils, or terrestrial groundwater.

The apparent formation of higher-molecular weight oxidized (and partly unsaturated) structures in the drier *Campina* system were by tendency reflected by a lower pH (~ 3.7; range 3.6 − 3.9) and older calibrated radiocarbon ages (~11 yrs); whereas slightly higher pH (~ 3.8; range 3.6 − 3.9, and incl. Rio Aşu with the maximal value 4.3) and younger radiocarbon age (~6 yrs) were reflected by

less oxidized, nitrogen-containing and lower-molecular weight formulae in the wetter Cuieiras site. This trend was even more pronounced in formulae indicative of plateau soils showing highest pH values in our study (~4.6, range 4.2 – 5.3), although we are missing radiocarbon data for these samples. It is particularly interesting that the same molecular trends in relation to pH were found in a study conducted in German forest soils (Roth et al., 2015), were pH was also strongly correlated with DOC concentration. It is however notable that their pH difference among samples was much higher (4.0 – 5.5) than in the samples used in our second PCoA analysis (Figure 4-5; pH range 3.6 - 3.9). Although the *envfit* analysis showed a clear gradient in pH, it was much less pronounced than in the study of Roth and coworkers, making it more probable that the observed differences in molecular composition and age of SPE-DOC relate to eco-hydrological properties besides pH, DOC concentration, or electrical conductivity (which were both strongly correlated to pH). This also demonstrated the value of the Amazon as a natural laboratory to disentangle pH effects from other variables that are usually correlated with pH (Roth et al., 2015; Rousk et al., 2010). However, the causes of such molecular trends in relation to pH remain unclear. Several authors have linked the release of specific high-molecular weight oxidized and unsaturated compounds to dissolved Fe and Al export, suggesting specific organo-mineral associations; this association has also been demonstrated by lab experiments (Coward et al., 2018; Galindo and Del Nero, 2014; Linkhorst et al., 2016; Raeke et al., 2017; Riedel et al., 2013). Active podsolization processes at Campina sites could lead to the remobilization of such indicative markers (Bardy et al., 2011). Also fungal enzymes have been suggested as agents of potential DOM transformation towards more higher-molecular weight, aromatic, and oxidized structures (Baldrian and Šnaijdr, 2011; Waggoner et al., 2015; Zavarzina et al., 2018). Campina forests usually show thick root mats above the actual soils, which could give rise to such fungal products. Acidic environments generally favor the presence of decomposer communities dominated by fungal species (Rousk et al., 2010), and dry conditions have been suggested as favorable, too (Bird and Torn, 2006).

Two minor observations of our study were linked to radiocarbon-correlated formulae and shifts in aromaticity between groundwater and streams. Formulae correlated to older age were generally smaller in terms of molecular weight (< 300-400 m/z), at least for the majority of CHO and CHNO formulae, whereas younger-age correlated formulae were of higher molecular weight (Figure A-12; Figure A-13) but not assigned to a specific category of formulae. This observation may reflect the initial decomposition of plant material in relatively young DOC in contrast to (preserved?) smaller breakdown products present in radiocarbon-older DOC but needs further investigation. Another minor observation of our study was a shift of DOM composition towards slightly less aromatic

stream DOM as compared to those found in piezometers of both sites (Figure 4-5, separation along coordinate 2). The variation was less pronounced than differences between sites and thus suggests similar processing that is however governed by site-specific factors as for example floristic composition (e.g., C/N ratios of litter, leaves, and root mats; Mc Clain et al., 1997) or decomposer community (Peyton Smith et al., 2015). As we lack compositional data from e.g., leaf litter leachates and microbial community, we cannot draw more conclusions from this observation. A general explanation for losses in aromaticity is light degradation. Water flowing in shallow streams allows for the penetration of light that is known to cause photo-degradation of DOM (Stubbins et al., 2010). Due to the small size of the streams, dark color, and a rather closed forest cover, this effect would be expectedly small, which agrees with the overall small difference (soils, 0.47, n=20 vs. streams 0.44, n=4).

4.3.3 Campina forests as a potential origin of specific Rio Negro markers

We found consistent overlap between a set of Campina markers with Rio Negro specific markers described by two independent, recent studies based on comparison of ultrahigh-resolution mass spectrometry (FTMS) data (Simon et al., 2019; Gonsior et al. 2016). The most robust range of overlap was limited to formulae with high aromaticity, low H/C, high O/C, and relatively high molecular weight (Figure 4-7, Figure A-14). This finding has great implications for the downstream transport of DOM in the Rio Negro watershed. First of all, it has to be noted that differences between the both reported Rio Negro-specific formula sets are evident (higher m/z range reported by Gonsior and coworkers) but this does not diminish the strong match in terms of O/C and H/C ranges (Figure 4-7a, c). The difference between instruments is less evident when all specific formulae are taken into account (i.e., when less robust, but still highly specific sets of Rio Negro-specific formulae are used; Figure A-14). The difference in highly robust signals can thus be explained by instrumental tuning and measurement conditions, an aspect that has been extensively discussed before (chapter 3; Simon et al., 2019). In general, it is likely that different FTMS instruments would yield different mass spectra of the same sample, which could, in theory, be minimized by parallel tuning (chapter 3; Hawkes et al., 2016). Differences in ion abundance patterns (spectrum envelope) will otherwise affect the outcome of multivariate statistics and ion-abundance-weighted averages of spectrum properties such as average values of e.g., *m/z* (mass center) or DBE (average number of double bond equivalents in a molecular formula). In most cases, different instruments will however capture the main gradients of variation in a sample set if differences are large enough (Hawkes et al., 2016). Thus, the consistency observed in our study is a highly encouraging result for future studies aiming

to compare dataset from different FTMS instruments working under different settings. We conclude that our study shows consistent overlap within the range of the most robust Rio Negro-specific markers reported by two independent groups, allowing for the successful matching of information generated by three different FTMS platforms.

It is long known that the wide distribution of whitesand areas in the Rio Negro basin is linked to its characteristic black waters (Goulding et al., 1988; McClain et al., 1997; Sioli, 1954). Bardy et al. (2011) reported on the formation of aromatic and carboxyl-rich WSOM in Amazonian Podzol profiles and established a link between WSOM residing in deep soil layers, groundwater, and streams based on similarities of specific UV absorbance, pH and electrical conductivity. In line with their findings, we here show that highly specific aromatic DOM compounds of the Campina site represent a potential molecular link between whitesand headwaters and the Rio Negro. Although Bardy and coworkers showed that Campina areas produce this marked DOM signal, other research showed that different sand-dominated systems (riparian corridors along the rivers) or wetlands contribute mainly to the DOC export of the Amazon and blackwater watersheds (Dosskey and Bertsch, 1994; McClain et al., 1997; Remington et al., 2007). It is now generally accepted that riparian corridors contribute mostly to the delivery of DOM and that these reservoirs are hardly depleted during storm events, making such singular events the main contributor of annual DOC exports (Laudon and Sponseller, 2018; Ledesma et al., 2015; Raymond and Saiers, 2010). In contrast to recent theory, we show here that a suite of highly specific markers differentiates a sandy, groundwater-influenced valley site classically defined as a riparian system from a rather upslope, dry Campina site dominated by precipitation (Zanchi et al., 2015), and that these markers allowed to directly link headwater processes to indicative downstream watershed signals. It was suggested earlier that this would be more probable in the Rio Negro catchment as compared to the whole Amazon watershed (McClain et al., 1997), as it is characterized by large amounts of upland soils classified as sand-dominated soils (Podzols, Arenosols) which give rise to high DOC exports (McClain et al., 1997; Bardy et al., 2011). Dark color and strong stratification of warm water bodies, partly due to stagnating flow connected to backwater effects (Meade et al., 1991), may also limit photochemical/ heterotrophic degradation (Goulding et al., 1988; McClain and Naiman, 2008). We provided the first direct molecular indication of this prediction, showing that remarkably dry, highly specific forest ecosystems leave a distinct imprint within the Rio Negro's exported DOM although contributing only secondarily to the fluxes of water. Annual discharge and DOC exports from both watersheds show a two times lower average annual stream discharge (based on the catchment area) but a two – three times higher annual DOC export from the Campina watershed (Zanchi et al., 2015;

Monteiro et al., 2014). This demonstrates that water and carbon cycles are decoupled on the molecular level and at the watershed scale because ecosystems are not behaving homogenously in neither their discharge nor carbon exportation behavior (Miguez-Macho et al., 2012a; Webb et al., 2018). In this specific region (the Rio Negro watershed), the riparian corridors may be sandy and do contribute mainly to the annual export of DOC (Remington et al., 2007; Miguez-Macho et al., 2012a), but our results show that they may not contribute to the specific signal that differentiates the watershed on the molecular DOM level. Campina systems seem to produce highly specific molecular signals for reasons that have to be clarified within future studies. Against first guess, the highly indicative set of Campina markers were not only found to overlap with Rio Negro samples from the proximity of its draining higher order river, the Rio Tarumã Açu (Figure 2-1; Simon et al., 2019) but also in samples upstream (sampled close to Novo Airão, 120 km northwest of Manaus) and thus much closer to the tributary (but still upstream) draining waters within the catchment of the Cuieiras site, the Rio Cuieiras. This implies that similar high-molecular weight oxidized aromatic compounds are exported upstream of Novo Airão, possibly by Campina systems as the one at Reserva Campina. It is not clear whether other Campina sites would show different compounds as might be suggested by the overall higher molecular weight of formulae described by Gonsior and coworkers (Gonsior et al., 2016).

Comparsion of datasets further allowed reconsideration of processes expected along aquatic flow paths. As stated earlier, formation of highly oxidized aromatics have been linked to the action of fungal enzymes (Waggoner et al., 2015; Waggoner & Hatcher 2017). Our results support the idea that such indicative products can originate from soils (especially in dry, acidic soils with fungal dominance; Bird & Torn 2006; Rousk 2010), rather than being produced through photochemistry in aquatic systems, which is reflected by the consistent decrease in aromatic structures in the DOM pool towards streams and rivers, as assessed from ion abundance information (Figure 4-10a – c; also for molecular formulae within the main range of Rio Negro-specific signals, i.e., H/C and O/C 0.4 – 0.8; Figure A-15). This decrease in aromatics was also surprsisingly consistent between the both Rio Negro studies, as opposed to e.g., predicted numbers of sulfur formulae; Figure 4-10d), and is also consistent with previous observations in soils and aquatic systems (Creed et al., 2015; Hutchins et al., 2017; Lambert et al., 2016). This stresses the importance of processing in lower-order streams (Hernes et al., 2017; Hutchins et al., 2017; Simon et al., 2019). The comparison of datasets revealed that nitrogen- and especially sulfur containing formulae clearly differentiated the three datasets considered in this study (Figure 4-8). Although both types of formulae may be affected by anoxic conditions, and also by changes in the connectivity of riparian systems (Boye et

al., 2017; Lynch et al., 2019; Peyton Smith et al., 2017), such differences may also be due to instrumental effects as heteroatom-containing formulae are harder to resolve and are often detected only at low ion abundance. Samples from a suite of channels separated by river islands (Figure 2-1; Gonsior et al. 2016), i.e., cluster b in Figure 4-9, were astonishingly similar to each other besides the large spatial extent covered, as compared to mainstem and tributary samples (cluster c; Simon et al. 2019) and headwaters (cluster a, this study) that were more dissimilar. This effect may point towards more homogenized DOM in aquatic systems as compared to soil environments (Kellerman et al., 2015; Lynch et al., 2019). The large gap between headwaters and river samples may indicate the extent of variation that can be expected in between these extreme stages of the river network (i.e., lowest and highest Strahler order). Further attempts to compare FTMS data between labs are needed to gain overarching insights into DOM processing and mixing along terrestrial flow paths that allow exclusion of instrumental effects (chapter 3).

4.4 Conclusion and outlook

We here provide the first analysis of the direct molecular link between two contrasting whitesand ecosystems and their downstream river, the Rio Negro, on the basis of their specific DOM fingerprints. Highly oxidized phenolic and aromatic structures of relatively high molecular weight (>300 Da) classifiable as "tannin-like" molecules emerge as global markers of acidic and DOC-rich ecosystems but their origin remains elusive. The specific conditions of the Rio Negro basin, especially strong leaching, deep weathering, tight cycling of nutrients, produce the earth's largest blackwater stream and a unique natural laboratory. These unique conditions allow to study the dynamics of DOM decomposition among soils and aquatic systems on a wide spatial gradient and seemingly "reduced speed" due to the fast export of large amounts of young DOM and their "delayed" decomposition in aquatic system characterized by low penetration of light, stagnating flow due to backwater effects, and stratification.

Molecular fingerprints of low-pH/ high-DOC ecosystems and nitrogen and sulfur-containing compounds show potential to reveal new insights into watershed decomposition dynamics; however, more efforts are needed to improve comparability among labs and thus draw overarching conclusion about DOM fate and information content. It is also of urgent need to develop new tools that allow the identification of indicative markers such as phenolics, N- and S-containing compounds to elucidate their origin and formation.

Whitesands and their link to blackwater DOM evolution in the Rio Negro basin

The characteristics of DOM from clayey environments remain to be studied at higher detail; theoretical considerations suggest that plateau soils might export a potentially old and highly reworked DOM signal that is overprinted by a major young DOM imprint by riparian and upland systems. The wide occurence of such soil environments (McClain et al., 1997) and their contribution to watershed and basin-wide DOC exports have to be addressed in future studies in order to reveal the potential export of dilute, persistent and old molecules to marine systems at the expense of large amounts of preferentially degraded young phenolic constituents. Our study highlights the importance of integrated watershed research at different scales that enables improved process understanding, from the scale of interconnected soil pores to river networks.

5 Identification of ecosystem markers by Orbitrap tandem mass spectrometry

5.1 Introduction

Dissolved organic matter (DOM) is defined as an ubiquitous, supramolecular, self-assembling and complex mixture of organic compounds that pass through a filter (mostly in the range 0.1 - 1 µm; Piccolo, 2001; Wells and Stretz, 2019; Zsolnay, 2003). The complex interactions of DOM and soil organic matter (SOM) form the basis of key ecosystem services (Bünemann et al., 2018; Delgado-Baquerizo et al., 2016; Kästner and Miltner, 2018; Kögel-Knabner, 2017; Lange et al., 2015; Sanderman et al., 2017). DOM thus depicts a fingerprint of ecosystem activities (Kaiser and Kalbitz, 2012; Kästner and Miltner, 2018; Roth et al., 2014). We still struggle to understand the conformation, composition and structure of the original DOM sample on the molecular level, despite advances in ultrahigh resolution mass spectrometry (FTMS; Hawkes et al., 2019; Malik et al., 2016; Roth et al., 2014; Simon et al., 2018). Novel molecular approaches to identify distinct markers in DOM are required to tackle these problems. Previous studies have argued that DOM found across aquatic gradients has been recycled up to a degree when source imprints have essentially vanished through universal patterns of degradation, which renders the DOM residues undistinguishable by means of recent ultrahigh resolution mass spectrometry (FTMS) and spectroscopic techniques (Hawkes et al., 2018a; Mentges et al., 2017; Zark and Dittmar, 2018). An important implication of this finding is that there must exist an early time point or period when a source imprint can still be detected by recent technology. We hypothesize that DOM residing in near-surface layers of soil, with close contact to plant inputs and microbial activity, would carry such a signature.

Analysis of structural detail in DOM is difficult due to the inherent complexity that arises from a multitude of processes in open and living systems (e.g., changing boundary conditions, sorption, rewetting, exudation, lysis), leading to inherently inseparable mixtures of thousands of individual constituents (Brown et al., 2016; Malik et al., 2016; Miltner et al., 2012). The latter problem also emerges in controlled systems that involve less diverse starting materials and single processing steps on short timescales, e.g. in lignin depolymerization or food processing (Kuhnert et al., 2013; Qi et al., 2016), but emerge also from spontaneous abiotic reactions of single compounds (Hemmler et al., 2017; Zherebker et al., 2015) or their rapid diversification through microbial metabolism (Lechtenfeld et al., 2015; Noriega-Ortega et al., 2019). As a consequence, many of the compounds

that are found in terrestrial and aquatic DOM resist usual attempts of separation, isolation and structure elucidation (Brown et al., 2016): DOM exerts a strong complexity on the nominal mass level (isobaric complexity) in ultrahigh resolution mass spectrometry data (FTMS) and is expected to host even more compounds on the isomer level (i.e., compounds with the same exact mass and molecular formula but differing structural arrangement; Brown et al., 2016; Hawkes et al., 2018; Zark and Dittmar, 2018).

Despite this isomeric ambiguity, it is a common practice to classify molecular formulae derived from exact mass data to structural domains in Van Krevelen space (D'Andrilli et al., 2015; Roth et al., 2014). The Van Krevelen plot depicts the chemical space formed by the stoichiometric ratios of hydrogen to carbon atoms (H/C) and oxygen to carbon atoms (O/C) in a molecular formula (Kim et al., 2003). The proximity of molecular formulae in this display is taken as a measure of chemical similarity and subsequently used to bin formulae into structural domains. Many authors have argued that structure annotation by molecular formula remains preliminary until proven by complementary methods, an objective impeded by the inherent complexity detailed above (Davies et al., 2015; Reemtsma, 2010; Rossel et al., 2016). When the use of a-priori defined domains is chemically imprecise or inconsistent it may also promote misconceptions (D'Andrilli et al., 2015). On top, present-day structural databases do cover only a minority of the molecular formulae encountered in DOM, typically allowing for the annotation of < 5% of formulae (Brown et al., 2016; Petras et al., 2017; Zhang et al., 2014). The interpretation of ultrahigh resolution mass data by abundance-weighted average indices is another usual practice (Boye et al., 2017; D'Andrilli et al., 2015; Koch and Dittmar, 2016; Mann et al., 2015; Mentges et al., 2017). While these parameters are effective for a broad characterization of samples or the comparison of larger sample sets, they obscure the molecular detail that the technology provides.

One way to peer into the chemical makeup of single formulae are fragmentation experiments that are readily available in most FTMS instruments. Relatively wide isolation windows (~ 1 Da) and poor chromatographic separation hinder the isolation and subsequent fragmentation of single molecular formulae (which could still represent isomeric mixtures), leading to so called "chimeric" spectra (Petras et al., 2017). Dedicated studies have shown that even with isolation of single masses by sophisticated technology not available to every lab, or chromatographic separation of isobars at hand, DOM fragmentation patterns seem to be rather universal as detailed above (Hawkes et al., 2018a; Witt et al., 2009; Zark et al., 2017; Zark and Dittmar, 2018). However, there are also authors that point out differences in fragmentation patterns after fractionation and separation by

multidimensional chromatography (Brown et al., 2016). What is common to all these studies is that they focus on the major product ion peaks (fragments) which usually make up $60 - 70\,\%$ of the total product ion abundance (Hawkes et al., 2018a; Zark and Dittmar, 2018). These major product ions relate to sequential losses of mainly CO_2, H_2O, CO and CH_4 units which agrees well with the required polarity of organic compounds soluble in water. These losses (in the following, termed mass differences, MD or MDs) do not allow a clear identification of a structural unit and thus are non-indicative and ambiguous identifiers (an overview of the associated MDs is presented in Table A-10 of the Appendix; Brown et al., 2016; Cortés-Francisco and Caixach, 2015; Witt et al., 2009; Zark and Dittmar, 2018). However, early studies (Capley et al., 2010; McIntyre et al., 2002; Nimmagadda and McRae, 2007; Perdue et al., 2007; These et al., 2004) have shown that there are recurring low m/z product ions (e.g., m/z 95, 97, 109, 111, 123, 125, 137, 139, 151 and 153) that might represent a diverse but limited set of core structural units being substituted with a limited set of functional groups, yet in different amounts and configurations. Later studies have followed up on this idea (Arakawa et al., 2017; Bell et al., 2015; Brown et al., 2016; Zherebker et al., 2015). It is intriguing that the recurring patterns in MS^2 data show parallel features of the original mass spectra (MS^1) in FTMS experiments. For example, the usual mass spacing of 2, 12 and 14 Da between main peaks are also found in the product ion spectra, however, only at lower m/z (Cortés-Francisco and Caixach, 2015; Stenson et al., 2003; These et al., 2004). The MD information encoded within these low-m/z MS^2 peaks has not yet played a major role in the debate about the isomeric complexity of DOM, whereas MS^1 data is close to become a routine analysis. Kunenkov and coworkers described a procedure to analyze the occurrence frequency of a mass spacing in FTMS data (MS^1) to reveal higher-order units inherent to the data structure, such as the 2, 12 or 14 Da mass spacing described before (Kunenkov et al., 2009). By applying their approach, the authors successfully revealed a novel inherent structural unit with the formula $C_7H_6O_4$ (154.027 Da), that they preliminary assigned to the structure of dihydroxylbenzoic acids. We would not expect to find such a "higher-order structural unit" in MS^2 data besides the major known ones (non-indicative MDs, Table S-1). We would rather assume that the proposed inherent isomeric diversification of DOM during its universal degradation favors the occurrence of common small losses. The probability of two molecules sharing a larger substructure (and thus, its MD) of exact same mass would be clearly reduced under the given assumption of a stochastic degradation process (implying that the range of realized structural space could be shifted by changing boundary conditions; i.e., in an incubation) which would make the occurrence of (specific) indicative MDs more favorable. This instance has been experimentally simulated by addition of known structures to the DOM mixture (Brown et al., 2016; Hawkes et al., 2018a).

Identification of ecosystem markers by Orbitrap tandem mass spectrometry

An approach to circumvent the problem of unknown substructure diversity is to hypothesize about potential structural units that would be present in DOM if there was a source imprint. For example, lignin-related compounds should indicate methoxy or methyl losses (Liu et al., 2011; Zark and Dittmar, 2018); glycosides should indicate the loss of a sugar unit and hydrolysable tannins would be expected to indicate losses of galloyl units (Engström et al., 2015; Gross, 2009). Flavon-3-ols and flavan-3-ols would be expected to show variable retrocyclisation products (Fabre et al., 2001; Galaverna et al., 2015; Miketova et al., 2000; Yuzuak et al., 2018). It has been recently shown that stepped fragmentation experiments are a way to separate mixtures of isomers (Dit Fouque et al., 2016). A consequence of the high isomeric diversity of DOM would also be that each resolved FTMS peak represents a mixture of ions that might respond differently to fragmentation, i.e., exerts a continuum of fragmentation sensitivities. Similarly, it is unclear whether the MDs between recurrent low-m/z product ions and single precursors could be taken as a rough measure of isomeric complexity. Several authors have described a confined area of formulae in the center of typical DOM peak distributions in Van Krevelen space which relates to most intense and ubiquitous peaks, and apparently also links to the most aged/ recycled DOC in terms of radiocarbon content (Kellerman et al., 2014; Kew et al., 2017; Lechtenfeld et al., 2014; Roth et al., 2014). Intriguingly, the respective region of the Van Krevelen plot has also been assigned to highest expected isomeric complexity as dictated by stoichiometry and valence (degrees of freedom; Hertkorn et al., 2008, 2007; Petras et al., 2017; Reemtsma, 2010). We thus would expect that most intense peaks in this "region" of chemical space would match more frequently to indicative MDs.
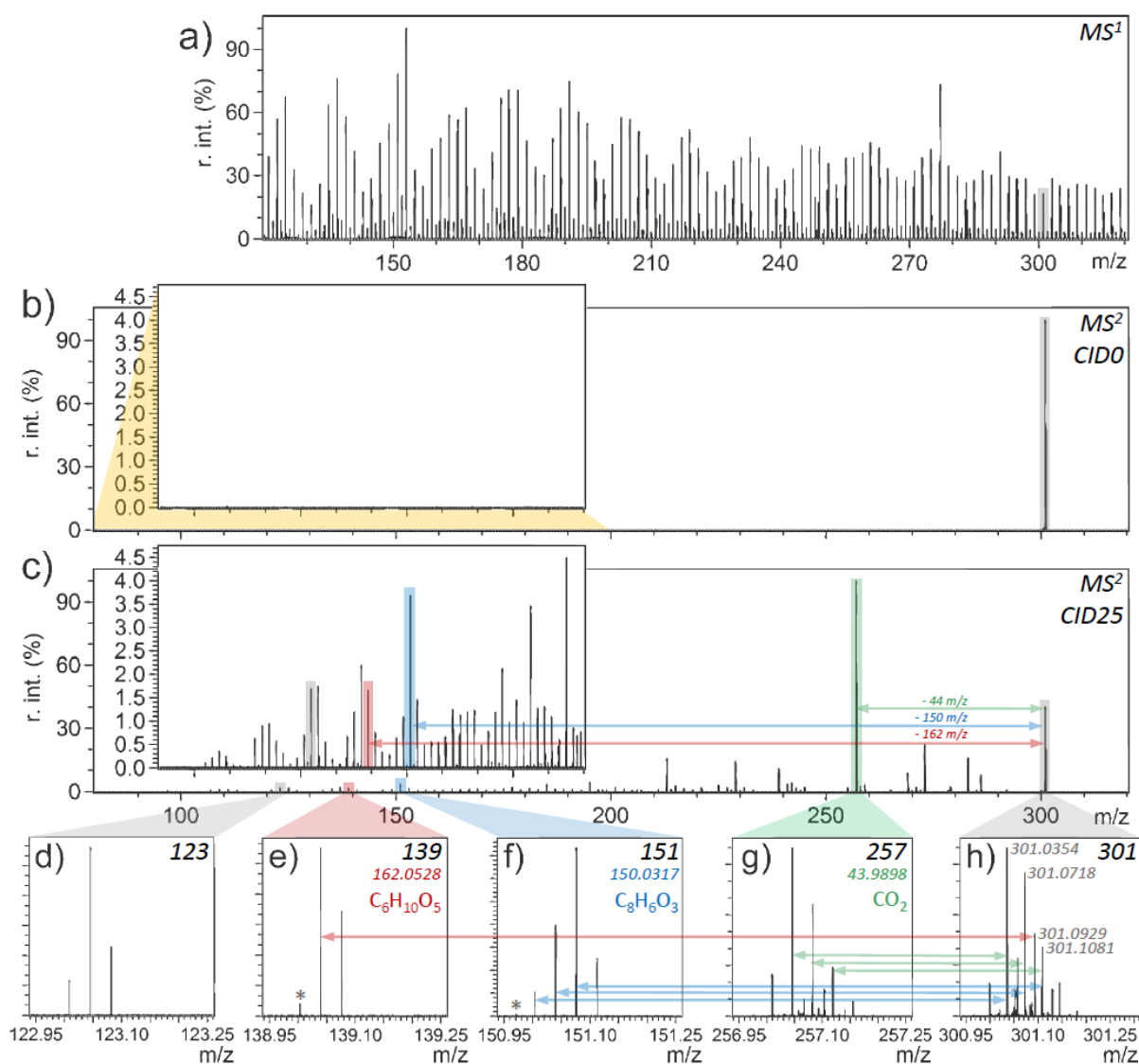
Figure 5-1. Orbitrap tandem MS data of DOM under different measurement modes. a) Detail of the initial MS^1 spectrum of the W1-5 sample. Note the lower limit of the scan range at m/z 120 (upper limit m/z 1000). b) Non-fragmented (CID 0) isolated isobar (isobaric precursor ion mixture, IPIM) at m/z 301 (see detail in panel h). No ions at other m/z values are detected. Inset shows the lower mass range below m/z 200, ~20fold enlarged. c) Tandem mass spectrum (MS^2) of IPIM at m/z 301 obtained at a normalized collision energy of 25 (CID25) and similar inset as in b). Panels d – h) Isobaric detail (exact mass) of four product ion clusters at CID25 (d – g) and the initial IPIM at m/z 301 (CID0, n ions=44). Four peaks assigned in h) were assigned the molecular formulae $C_{15}H_{10}O_7$, $C_{16}H_{14}O_6$, $C_{13}H_{18}O_8$ and $C_{17}H_{18}O_5$ (in order of increasing exact m/z). For those ions, neutral losses are indicated by the arrow between isobars (301/ 257, green; 301/ 151, blue, and 301/ 139, red). The nominal MDs of 44, 150 and 162 (panel b) can be assigned to exact MDs of product ions (panels d – g), such as neutral losses of CO_2 (a common, non-indicative MD, 3 out of 27 matches to IPIM at m/z 301 shown), $C_8H_6O_3$ (an indicative loss equivalent to a retrocyclisation loss from flavonol-type-molecules, 3/ 4 matches shown) and $C_6H_{10}O_5$ (indicative MD equivalent to neutral loss of glucose loss from flavonoids, 1/ 2 matches shown). Product ions at m/z 123 (d) had absolute intensities (ion abundances) of 20, 40 and 90, being equivalent to signal-to-noise ratios of ~ 7, 13 and 30; the signals were stable in time and detected in repeated measurements. Exemplary peaks that were considered noise are marked with an asterisk (*) in panels e and f.

Identification of ecosystem markers by Orbitrap tandem mass spectrometry

To gain better understanding of the ecosystem information encrypted in chimeric fragmentation spectra of DOM and its behavior upon fragmentation, we selected a set of pure compounds that covered a range of structural units and molecular sizes and compared their indicative MDs to similar data obtained from selected isobaric precursor ion masses ("IPIMs"; m/z 241, 301, 361, 417) of a soil DOM sample. This sample had been part of preliminary studies (Benk et al., 2018; Roth et al., 2014; Simon et al., 2018). We then evaluated the similarity of DOM precursors in terms of their indicative MDs and checked a natural product database for consistency of structural suggestions. Finally, we analyzed whether individual DOM precursor formulae exerted different fragmentation sensitivities and matching frequencies and how this was related to precursor properties.

5.2   Results and Discussion

5.2.1   Evaluation of the MD matching approach

Our analysis of MDs in DOM was highly congruent with previous observations, showing ubiquitous losses of small non-indicative oxygen-containing functionalities (Table A-10) while also revealing often-overlooked detail present within the data (Figure 5-1, Table 5-1; Hawkes et al., 2018; Osterholz et al., 2015; Witt et al., 2009; Zark and Dittmar, 2018). Multiple additional observations support the revealing of inherent structural information by our approach. One of those comes from the assigned molecular formulae and especially their predicted heteroatom content (O, N, S). As expected, matching MDs were constrained by their precursor formula and vice versa. Formulae rich in oxygen were able to expel more oxygen-containing MDs than oxygen-poor formulae that tended to lose CH_2 or $CH_3^\bullet$ (and CO) units instead. Most notably, no formulae matched to an MD that would have exceeded the number of atoms actually present in the assigned molecular formula, a condition that has not always been met in such data analyses in the past (due to low-res tandem MS; Capley et al., 2010). Sulfur- and Nitrogen-containing formulae – and only those – indicated the release of previously described element-specific MDs (Pohlabeln and Dittmar, 2015; Wagner et al., 2015; Zhang et al., 2014). We furthermore did not observe an increase in the amount of false positive matches upon widening of the tolerance window applied during the MD matching process (Figure A-17, increase from 2 to 10 ppm, at 200 m/z). Lastly, formulae resisting fragmentation did not match to any MD (Figure A-18), whereas molecular formulae that were fragmented to near completeness showed a wide range of matches.

Identification of ecosystem markers by Orbitrap tandem mass spectrometry

Table 5-1. List of 38 indicative mass differences (MDs) from the standard compound set that were also found in DOM. Creosol and vanillic acid were added (marked by #$^{)}$) as neutral molecules although not part of the substance-based MD list. Prec. = precursor. Compound ID refers to Figure 2-3. Contribution of MS^3 data is marked with an asterisk. IDs are put in brackets if the MD was detected below 1% base peak intensity across three CID levels. MDs that contributed only below <1% base peak intensity were taken into account only if detected in at least two substances. Occurrence, matches across 159 precursor formulae.

Formula	Exact MD	From comp ID.	Explanation	Occurence
$CH_3^\bullet$	15.0235	1, 4, 5, 6, 6*	Methyl radical, loss from radical ion on (6)	19 (12.0%)
$CH_2O_2^\bullet$	44.9977	(2), (8)	Formic acid equivalent, radical	2 (1.3%)
CH_2O_2	46.0055	(6*), 13, (13*)	Formic acid equivalent	20 (12.6%)
C_3H_6O	58.0419	10	Acetone equivalent; comb. C_2H_2O (ethenone) + CH_4 (Sirius)	14 (8.8%)
$C_2H_4O_2^\bullet$	59.0133	1, (10)	Acetic acid, radical	10 (6.3%)
$C_4H_4O_2$	84.0211	10	Combination, C_3O_2 (carbon suboxide) + CH_4 (Sirius)	6 (3.8%)
$C_3H_2O_3$	86.0004	(1), 10	Combination, C_3O_2 (carbon suboxide) + H_2O (Sirius)	13 (8.2%)
$C_2H_2O_4$	89.9953	(10), 13*	Oxalic acid equivalent	20 (12.6%)
C_3O_4	99.9797	8	Combination, CO_2 + 2x CO	15 (9.4%)
$C_4H_6O_3$	102.0317	10	Combination, C_3O_2 + CH_4 + H_2O (Sirius)	13 (8.2%)
$C_4H_8O_3$	104.0473	14	Hydroxybutyric acid equivalent	16 (10.1%)
$C_6H_4O_2$	108.0211	12*, 13*	Benzoquinone equivalent	7 (4.4%)
$C_6H_6O_2$	110.0368	10	Benzenediol equivalent; comb., C_3O_2 + CH_4 + C_2H_2 (Sirius)	2 (1.3%)
$C_4H_2O_4$	113.9953	(8), (10)	Butynedioic acid equivalent	9 (5.7%)
$C_4H_8O_4$	120.0423	13	Tetrose equivalent	15 (9.4%)
$C_7H_6O_2$	122.0368	10, 12*, 13*	Loss from flavonols; Comb. on (10): C_3O_2 + C_4H_6 (Sirius)	2 (1.3%)
$C_7H_8O_2$	124.0524	10, Prec. (5)	3-Methoxyphenol (m-Guaiacol) unit	12 (7.5%)
$C_6H_6O_3$	126.0317	(10), 11, 14	Phloroglucinol unit	11 (6.9%)
$C_5H_4O_4$	128.0110	10	Comb., C3H4O2 + C2O2 (Sirius)	8 (5.0%)
$C_7H_6O_3$	138.0317	10, 11, (13*)	Comb., C6H6O2 + CO (Sirius)	7 (4.4%)
$C_8H_{10}O_2$	138.0681	Prec. (4)	#Creosol unit	7 (4.4%)
$C_6H_{10}O_4$	146.0579	14	Sugar unit, Rhamnose	8 (5.0%)
$C_6H_{12}O_4^\bullet$	147.0657	14	Sugar unit, Rhamnose, radical form	1 (0.6%)
$C_8H_6O_3$	150.0317	12*, 13*	Specific loss from flavonols	16 (10.1%)
$C_7H_4O_4$	152.0110	9, 11	Incomplete gallic acid unit; H2O retained at precursor	7 (4.4%)
$C_9H_6O_3$	162.0317	7	Caffeoyl unit	9 (5.7%)
$C_6H_{10}O_5$	162.0528	12, 13	Sugar unit, Glucose	9 (5.7%)
$C_9H_8O_3$	164.0473	9, 10, Prec. (2)	p-coumaric acid; Comb. on (10): C7H6O3 + C2H2 (Sirius)	7 (4.4%)
$C_8H_8O_4$	168.0423	Prec. (1)	# Vanillic acid unit	6 (3.8%)
$C_7H_6O_5$	170.0215	9, 11, Prec. (3)	Gallic acid unit	5 (3.1%)
$C_7H_{10}O_5$	174.0528	7, (14)	Quinic acid unit (7)	7 (4.4%)
$C_8H_4O_5$	180.0059	12*, 13*	Loss from flavonols	4 (2.5%)
$C_9H_8O_4$	180.0423	(7), 10	Caffeic acid unit; Comb. on (10): C7H8O2 + 2x CO (Sirius)	4 (2.5%)
$C_8H_6O_5$	182.0215	11	Comb., C6H6O3 (e.g., Phloroglucinol) + C2O2 (Sirius)	4 (2.5%)
$C_9H_{10}O_4$	182.0579	(7), (9)	Comb. on (9): Coumaryl loss + 2x H2O (Sirius)	5 (3.1%)
$C_7H_8O_6$	188.0321	(9), 11	Comb., C7H6O5 (e.g., Gallic acid) + H2O (Sirius)	4 (2.5%)
$C_9H_6O_5$	194.0215	12*, 13*	Loss from flavonols	6 (3.8%)
$C_{13}H_{12}O_6$	264.0634	11	Degrad. central Catechin C ring after A or B-ring abstraction	2 (1.3%)

Table 5-2. List of 8 non-indicative mass differences (MDs) extracted from the standard compound dataset also reported for DOM (Table A-10). Contribution of MS3 data is marked with an asterisk at the compound number (numbers refer to Figure 2-3). Compound identifiers are put in brackets if the MD was detected below 1% relative intensity (based on base peak) across three CID levels. MDs that contributed only below <1% were only taken into account if detected in more than one compound. Occurrence, matches across 159 precursor formulae.

Formula	Exact MD	From comp # No.	Explanation	Occurence
H_2O	18.0106	(2), 10, 13*, (14)	Water	64 (40.3%)
CO	27.9949	(4), (6*), (8), 12*, 13*	Formyl transf./ Carbon Monooxide	44 (27.7%)
C_2H_4	28.0313	4, 5	β-oxidation/ fatty acid synthesis	44 (27.7%)
C_2H_2O	42.0106	(2), (4), 6*, 8	Hydroxypyruvic acid/ $-H_2O$	13 (8.2%)
CO_2	43.9898	1, 2, 3, (7), 8, 10, 11, 12*, 13*	Carbon dioxide/ Carboxyl group	96 (60.4%)
CH_2O_3	62.0004	10, 13*	Comb., $CO_2 + H_2O$	43 (27.0%)
C_2O_3	71.9847	(1), 8, (10), 12*, 13*	Comb., $CO_2 + CO$	31 (19.5%)
C_3O_5	115.9746	8	Comb., 2x $CO_2 + CO$	18 (11.3%)

The combination of these observations leads us to the conclusion that the mass difference matching approach resolves molecular detail of a true biogeochemical signal and does not yield random matches, although pairwise mass difference calculation between precursors and products would allow for this. A random matching result to MDs of seemingly wrong precursor compositions (e.g., loss of sulfate from a sulfur-free precursor; four CO_2 losses from a precursor with only seven oxygen atoms) would be expected if the calculated MDs were derived from noise and not from an inherently structured biogeochemical signal. This is a notable finding as it suggests that it may be possible to deconvolute isobaric precursors (chimeric mass spectra) in the future.

5.2.2 DOM ecosystems imprints revealed

The standard compound set (Figure 2-3) occupied a wide range of H/C and O/C ratios in accordance with other known members of structurally related classes of compounds (Figure 5-2a). Although these classes of molecules show a high degree of structural variability, their distribution in Van Krevelen space indicates a high degree of overlap, even when plotted against their molecular mass (m/z in Figure 5-2b). Although the group of tannins is quite well separated, close to none of their members is found in the assumed tannin region (Figure 5-2a), e.g. the both base structures of hydrolysable tannins, Ellagic acid (**#8**, $C_{14}H_6O_8$) and Gallic acid (**#3**, $C_7H_6O_5$). Condensed tannin structures (Catechin-based, **#10**) are a type of flavonoid that plots in the center of the diagram. Most tannic formulae do show a high molecular mass (> 800 Da), a feature usually not encountered in FTMS DOM data. This misfit depicts well the limitation of Van Krevelen diagrams to identify structural detail in DOM (Davies et al., 2015; Reemtsma, 2010).
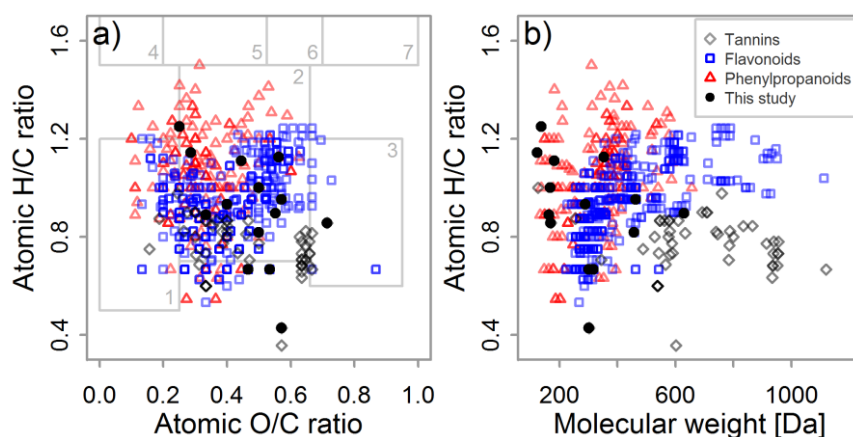
Figure 5-2. Limitations of Van Krevelen-bases interpretation. a) Van Krevelen plot of structurally different groups of substances from the KEGG database (tannins, n = 55; flavonoids, n=452; phenylpropanoids, n = 185). Grey boxes indicate regions of structural units taken from Minor et al., 2014; 1 – Condensed hydrocarbons, 2 – Lignin and/ or CRAM, 3 – Tannins, 4 – Lipids, 5 – Proteins, 6 – Aminosugars, 7 – Carbohydrates. b) Same data as in panel a, atomic O/C ratio replaced by molecular weight (Da).

The 14 pure substances expectedly showed a wide variability in their fragmentation behavior (Figure 5-3, Table A-7). The three small carboxyphenols (Figure 5-3, group A, Vanillic acid, Hydroxycinammic acid, Gallic acid) were characterized by a dominant CO_2 loss. Vanillic acid (**#1**) shared with members of group B (methoxyphenols and methoxy-quinones) the presence of a methoxy group which gave rise to the loss of a methyl radical ($CH_3^{\bullet}$, Table 5-1). This loss was the main common MD in group B (Creosol, m-Guaiacol, 2,3-Dimethoxy-5-methyl-1,4-benzo-quinone). The group of linked carboxyphenols (**#7 – #9**) was characterized by a variety of indicative losses and fragments, mainly connected to losses of whole substructures by cleavage of ester bonds (e.g., quinoyl and caffeoyl moieties from **#7**, Table 5-1). The structure of ellagic acid (**#8**) features two intramolecular lactone functionalities. In contrast to the ester bonds in substances **#7** and **#9**, these lactone bonds were exceptionally stable upon fragmentation and did only yield rich product spectra at higher CID energies (> CID25). Substance **#11**, containing a flavan-3-ol subunit, resembled especially substance **#9** through the presence of a gallic subunit that produced similar MDs ($C_7H_4O_4$, $C_7H_6O_5$, $C_7H_8O_6$) that represent an incomplete galloyl loss with retention of H_2O, a galloyl loss, or a combined galloyl and H_2O loss, respectively (Table 5-1). The preferential degradation of the galloyl moiety prevented a further fragmentation of the catechin structure. Both compounds shared the losses of $C_6H_6O_3$, but probably for different reasons (unmodified A ring in **#10**, abstraction of trihydroxybenzene from gallic acid unit in **#11**; Miketova et al., 2000; Poon, 1998).
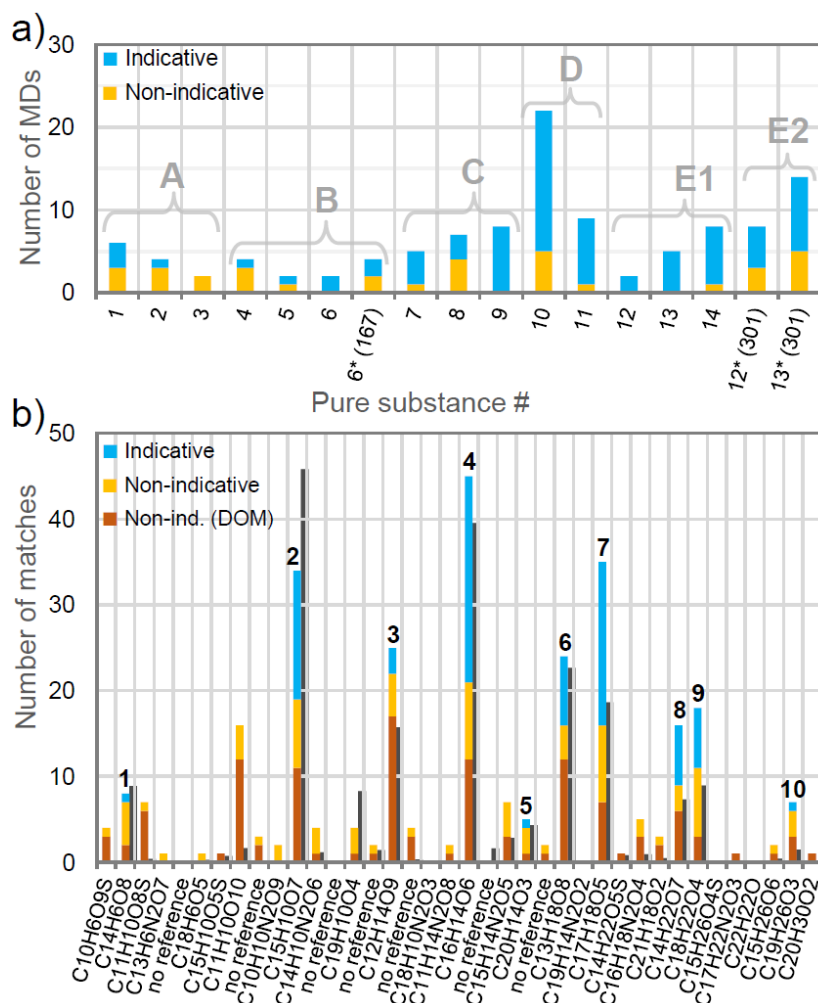
Figure 5-3. Tandem MS data from a) standard compounds and b) DOM. a) Shown is the total number of MDs, colors denote the absolute fraction of MDs that were also encountered in DOM samples (Table A-9; Table A-10). Groups are A) Small carboxyphenols, B) Small methoxyphenols and methoxyquinones, C) Linked carboxyphenols, D) Flavanol-related structures, and E) Flavonol glycosides and aglycones. b) Total number and fractions of MD matches of IPIM 301 (at CID25, n = 38, m/z increases to the right) to the lists of standard compounds (blue fraction, Table 1), non-indicative MDs (yellow, Table 2), and literature-known DOM MDs (orange, Table S-7). Black bars in the back show (dimensionless) initial intensity of the respective precursor for comparison. Only ten precursors contribute to the indicative matches (m/z 241, 11; m/z 361, 12; m/z 417, 13). Numbers indicate precursor formulae that are discussed in the text and other figures.

In general, Catechin (#**10**) had the most diverse product spectrum among all compounds investigated (Figure 5-3a, group D) and showed some indicative MDs connected to the initiation of retrocyclisation reactions (fragments at *m/z* 205, 203, 179, 151, 125, and 109, Table A-7; Galaverna et al., 2015; Rockenbach et al., 2012; Yuzuak et al., 2018). The group of flavonoids (substances **12** – **14**, containing flavon-3-ol cores, Spiraeoside, Isoquercetin, Myricitrin) showed a clear loss of the attached glycosidic sugar as the main MD (Table 5-1), differing only in the type of sugar (**12** and **13**, glucose, **14**, rhamnose). Moreover, this loss occurred in different ways in the three structures. Substance **12** fragmentation produced mainly the ion form of the aglycon, whereas the trend shifted towards the radical anion form in substance **14,** and being roughly even in **13**. This effect that has

106

been attributed to the exact location of the glycosylation site (Engström et al., 2015). It apparently also had an influence on further fragmentation of the aglycon, which proceeded in substance **14** (less so in **13**) and not at all in **12** (Figure 5-3a, E1). The aglycon ions at *m/z* 301 were further fragmented via MS[3] and yielded indicative retrocyclisation products (at *m/z* 179, 151, 121 and 107)(Fabre et al., 2001) that differed in their exact mass from those found in Catechin (#**10**, especially important in *m/z* 179 and 151).
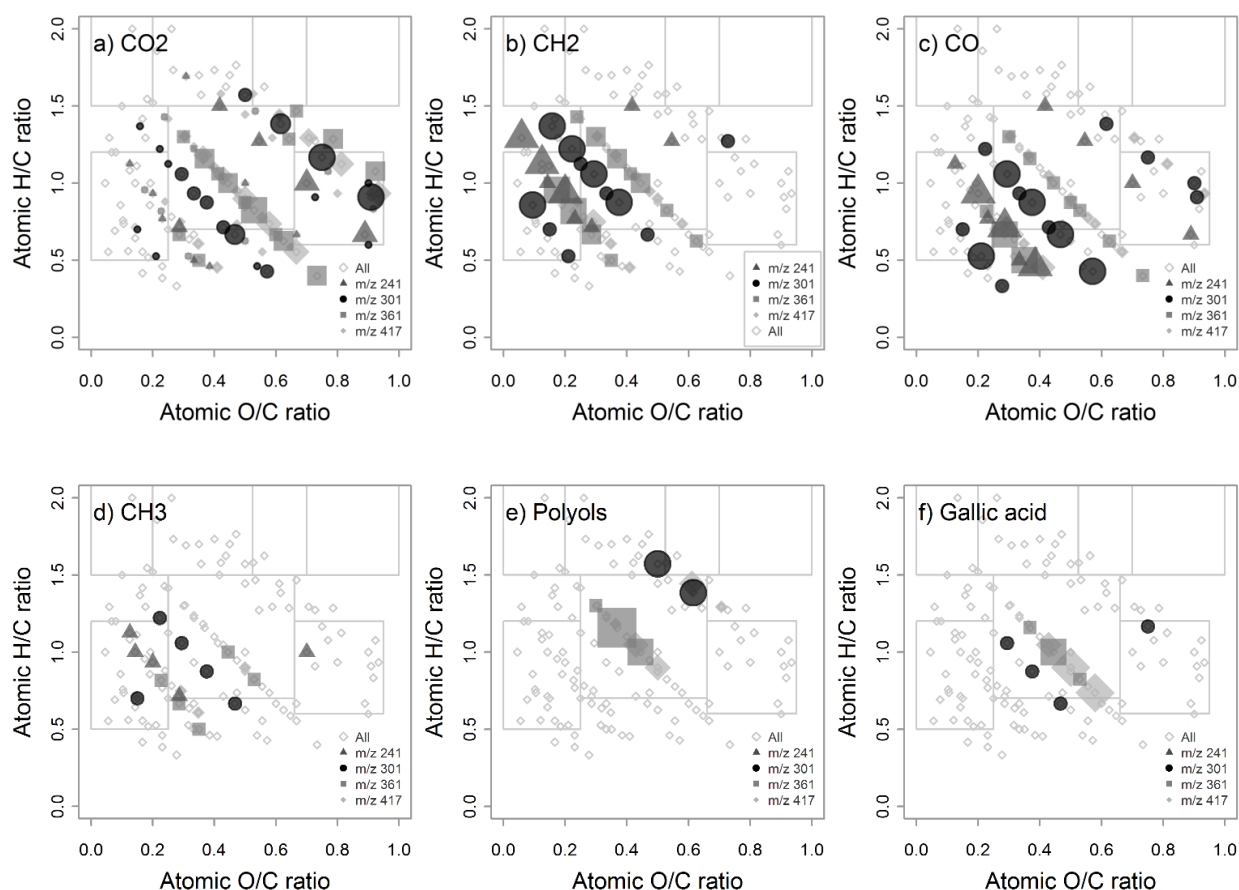
Figure 5-4. MD matching seen in Van Krevelen space. The plots show precursors with an assigned molecular formula at their atomic ratio of O/C and H/C. The number of matches to non-indicative (upper panels a – c) and indicative MDs (lower panels d – f) is encoded by the size of the symbol. Symbol and color are encoding for one of four IPIMs (see legend). Small open diamonds show all precursors with an assigned molecular formula (n=127). IPIM at m/z 301 is highlighted by black dots. MDs shown are a) CO_2 (max=4, symbol size reduced by factor 2), b) CH_2 (max=4, symbol size reduced by factor 2), c) CO (max=2), d) Methyl radical (max=1) e) MDs equivalent to polyol losses (max=3), and f) MDs equivalent to gallic-acid related losses (max=3).

The analysis yielded non-indicative and indicative MD patterns that were compared to MDs encountered in DOM (Figure 5-3b, Figure A-19a–c). As expected from the literature, non-indicative MDs dominated the DOM tandem MS data and were also found in standard compound data (Table 5-2). For example, CO_2 losses were observed in nine structures. This loss was not limited to structures with attached carboxyl functionalities (as in substances #**1-3** and in the set of small

organic acids fragmented by Zark et al., 2017). Ring cleavage and rearrangement reactions, most probably from neighboring hydroxyl and/ or carbonyl/ keto functionalities, also produced a neutral loss of CO_2 and did so at similarly low collision energies. For example, we observed CO_2 losses in flavonoid aglycons (spiraeoside **#12***, and quercetin **#13***, but not in myricetin **#14***, MS[3] data not shown) or catechin (**#10**), and to a lower degree also in ellagic acid (**#8**, probably originating from the lactone functionality; Capley et al., 2010; Fabre et al., 2001; Miketova et al., 2000; Mullen et al., 2003; Witt et al., 2009; Wyrepkowski et al., 2014). Besides a general dominance of non-indicative MDs (Table A-10), DOM product ion spectra also showed remarkably diverse features in the lower *m/z* range. We repeatedly detected dominant ions at *m/z* 95, 97, 109, 111, 123, 125, 137, 139, 151 and 153, in part, these product ions were also detected in the standard compounds.

Among the most prominent indicative features was the methyl radical loss from methoxylated compounds (Capley et al., 2010; Liu et al., 2011; Zark and Dittmar, 2018), a feature also detected in 19 oxygen-poor DOM precursors (on average, O/C = 0.3, Figure 5-4d). The indicative loss of methyl radicals from methoxy functionalities of aromatic ring systems is an expected diagnostic MD (Liu et al., 2011; Zark and Dittmar, 2018) for lignin due to its methoxylated monolignol building blocks (coniferyl, sinapyl alcohol) but evidence for such MDs in DOM has just been reported recently (Brown et al., 2016). The presence of methoxy functionalities in soil DOM, which carries a mixed signal of fresh and freshly decomposing plant material and more decomposed soil organic matter (SOM), is not surprising and reflects the high potential for transformation of non-soluble plant material by the decomposer community (Liu et al., 2011; Malik and Gleixner, 2013; Waggoner et al., 2017). The loss of a methyl radical was paralleled by similar van-Krevelen patterns of CH_2 and CO losses, all of which were found to be more prominent in similarly oxygen-poor formulae (Figure 5-4b, c), in line with previous reports (Brown et al., 2016). Radical losses from even-electron formulae were also observed in DOM before (Capley et al., 2010). Although we found matches to previously described CH_2O MDs in DOM (indicative of a methoxy group), the standard compounds did not indicate this loss; maybe due to their small size. Larger molecular scaffolds as those proposed by Liu and coworkers (Liu et al., 2011) might show a different behavior. We also found evidence for losses of methanol moieties. The presence of methanol losses has been considered an artifact due to potential methyl ester formation between carboxyl functionalities and methanol used for solid-phase extraction (SPE; Flerus et al., 2011). The DOM sample used herein was freshly extracted and thus not stored for long time (< 2 weeks at -20°C). Our group was also able to show recently that no dilution of the ^{14}C signal by radiocarbon-dead methanol occurred during extraction and storage of the same initial sample (Benk et al., 2018).
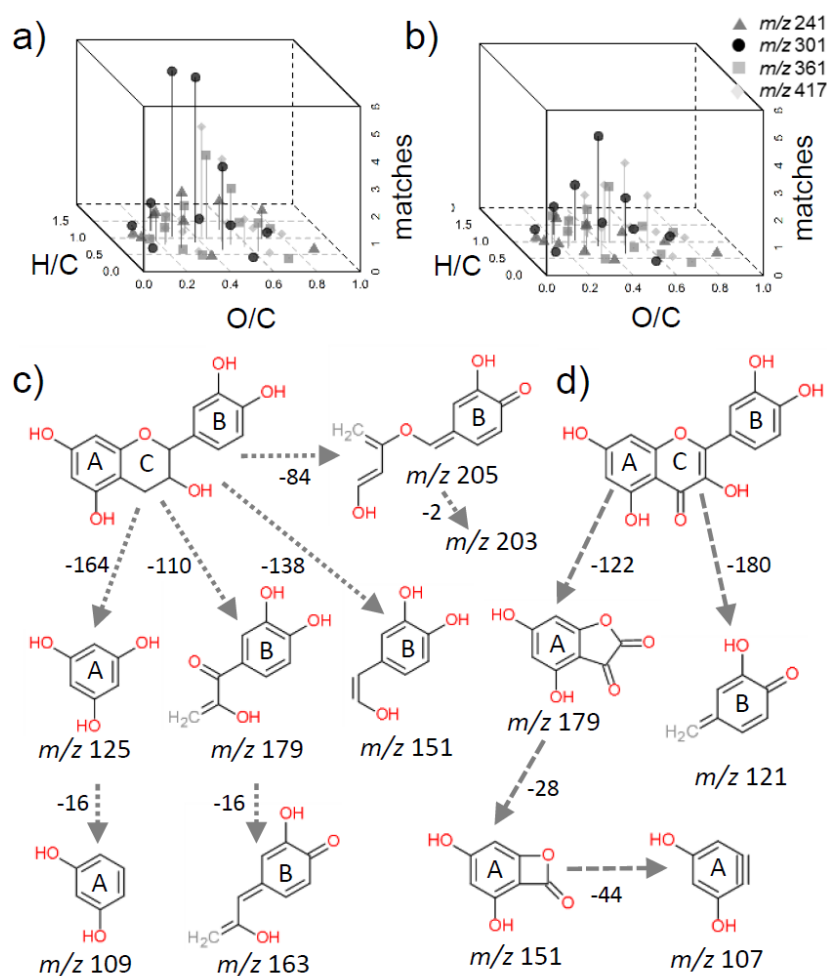
Figure 5-5. Flavonol and flavanol-related MD matches in DOM. 3D-Van Krevelen plots show matches against a) six MDs indicative of flavan-3-ol scaffolds and b) four MDs indicative of flavon-3-ol scaffolds. c) Scheme showing the major neutral precursors and products of the suggested fragmentation pathway of a flavan-3-ol (shown is catechin, #10). Related MDs are given as nominal m/z (Table 5-1; Galaverna et al., 2015; Miketova et al., 2000; Yuzuak et al., 2018) d) Similar scheme of the suggested fragmentation pathway for a flavon-3-ol (shown is quercetin, core structure in #13; Fabre et al., 2001).

Ester-linked carboxylated phenols (**#7**, **#11**) and O-glycosides (**#9**, **#12**, **#13**, **#14**) all indicated the cleavage of their central O-link by relatively low energies, leading to the detection of MDs linked to hydrogen-rich substructures (Engström et al., 2015; Leenheer et al., 2001; Witt et al., 2009).

Despite the widely accepted dominance of oxygen-rich losses in DOM, it is difficult to assign these MDs and their assigned molecular formulae to combinations of either CO, H_2O or CO_2, as the number of oxygen atoms in such higher-order MDs is relatively low compared to the number of carbon and hydrogen atoms – much similar to the $C_7H_6O_4$ unit that was described by Kunenkov and coworkers (equivalent to the exact MD of dihydroxylbenzoic acids; Kunenkov et al., 2009). For example, in MDs equivalent to losses of $C_8H_6O_3$ and $C_6H_{10}O_5$, which were observed in our study

(Figure 5-1e, f), C- or CH-rich losses would remain if all oxygen was assigned to simple losses of H_2O, CO_2 or CO.

Polyol losses from flavonoid glycosides and chlorogenic acid were a major indicative feature (Engström et al., 2015; Fabre et al., 2001; Ncube et al., 2014) and their equivalents were found 25 times in the four IPIMs (Figure 5-4e). Single precursors indicated losses of all three polyols taken into account here (glucose, rhamnose, quinic acid; e.g. DOM precursors $C_{13}H_{18}O_8$ and $C_{14}H_{22}O_7$ at m/z 301). We also observed matching to MDs equivalent to members of two series of highly indicative ring cleavage sequences (Figure 5-5) of flavon-3-ol (i.e., flavonoid aglycones; Engström et al., 2015; Fabre et al., 2001; 28 matches in total) and flavan-3-ol (i.e., catechin; Galaverna et al., 2015; Miketova et al., 2000; Yuzuak et al., 2018; 50 matches) structures. The molecular formula $C_{16}H_{14}O_6$ at nominal mass 301 showed strong matching to both series, suggesting the contribution of both structural families to that isomeric mixture. The use of a whole series of related MDs can, theoretically, improve confidence in structure assignment. Similarly, the use of more than one polyol MD can improve confidence in class annotation (i.e., "polyol-expelling formulae"). However, it has to be taken into account that chimeric tandem mass spectra hinder direct structure annotation (Petras et al., 2017). Our findings can thus be summarized as follows: We found multiple matches of potential higher-order MDs in a low-degraded DOM sample from upper forest soil supporting our hypothesis that such detail would be visible in such sample types. Yet we acknowledge that the low ion abundance of these signals (Figure 5-2) supports the general view that the molecular imprint of major plant biochemicals vanishes rapidly during initial decay. However, the high diversity of low-m/z fragments outweighs the high abundance of single product ions connected to common but non-indicative MDs such as CO_2, and reveals much more potential structural detail.

Differences between the fragmentation behavior of standard compounds and DOM have been widely described in the literature (Hawkes et al., 2018a; Leenheer et al., 2001; Novotny et al., 2014; Witt et al., 2009; Zark et al., 2017). As pointed out in the introduction, this observation has recently led to the consideration of intrinsic averaging as a main cause behind the high apparent similarity in ultrahigh resolution MS^1 and MS^2 features found across aquatic gradients (Zark et al., 2017; Zark and Dittmar, 2018), even after further reduction of complexity by chromatography (Brown et al., 2016; Hawkes et al., 2018a). The contribution of indicative MDs to the MD spectrum of different DOM precursors in our study adds important information to this debate. In the most recent literature, focus has been mainly put on common MD (ubiquitous) signals representing $60 - 70\%$ of fragment intensity (Hawkes et al., 2018a; Zark and Dittmar, 2018). Earlier studies have been limited to these

signals due to instrumental limitations, e.g. by low resolution tandem MS or low sensitivities in the lower *m/z* range. However, novel instrumentation, improved fragmentation cells, the possibility to measure product ion mass spectra at ultrahigh resolution, better comparability among instruments, and the use of heavy collision gas such as N_2 all allow an improved evaluation of low-*m/z* fragment signals (Bayat et al., 2018; Brown et al., 2016; Capley et al., 2010; Denisov et al., 2012; Dit Fouque et al., 2016; Ha et al., 2017; Snyder et al., 2016). Furthermore, Orbitrap instruments have recently shown better reproducibility in DOM (MS^1) measurements (Hawkes et al., 2016; Simon et al., 2018), which could also help to overcome problems of differentiating samples based on their MS^2 spectra (Zark and Dittmar, 2018). DOM molecular diversity is affected differently by continuous degradation along the aquatic continuum, causing simultaneous decreases in molecular features of MS^1 data and increases in isomeric complexity as deduced from MS^2 data. Both trends ultimately lead to undistinguishable features at both levels of MS data (Hertkorn et al., 2008; Mentges et al., 2017; Zark and Dittmar, 2018), making indicative MDs such as the methyl radical loss disappear (Hawkes et al., 2018a; Zark and Dittmar, 2018). The fact that we find such indicative MDs in supposedly less degraded soil DOM supports the theory of continuous DOM degradation and isomeric diversification in the aquatic continuum. It also puts special emphasis on soils as highly dynamic living reactors and their role as drivers of organic matter inputs into aquatic systems.

We interpret the existence of indicative higher-order MDs in DOM as a sign of a remaining source imprint of primary or recycled remains from plants, soil animals and microbes (Kästner and Miltner, 2018; Kögel-Knabner, 2017; Soucémarianadin et al., 2017; Waggoner et al., 2017). Recent studies indicate that precursor and product ions from specific molecular structures (added standard compounds) produce clear MS^1 and MS^2 signals in DOM elution profiles, despite the fact that many other molecules are present at the moment of measurement (Brown et al., 2016; Hawkes et al., 2018a). We thus assume that indicative product ion signals are not suppressed. The existence of low-*m/z* fragments that resemble the repetitive patterns found in MS^1 data suggests links between common low-*m/z* fragments and indicative MDs we found in this study (m/z 108, $C_6H_4O_2$, equivalent to a benzoquinone loss; m/z 110, equivalent to benzenediol loss, $C_6H_6O_2$; m/z 126, $C_7H_8O_2$, equivalent to a methoxyphenol loss; m/z 128, $C_6H_6O_3$, equivalent to a benzenetriol loss; Capley et al., 2010; McIntyre et al., 2002; These et al., 2004; Zark et al., 2017). The low-*m/z* region in tandem MS data of DOM is typically characterized by a multitude of signals at nearly every nominal mass (Figure 5-1b; Brown et al., 2016; Capley et al., 2010; Hawkes et al., 2018). This property is enigmatic but may also render assignments to specific MDs ambiguous, same as their detection in DOM. This work represents a first step in revealing ecosystem information encrypted

within chimeric DOM signals. Further understanding will benefit by complementary analyses based on extended sets of samples, standard compounds, and state-of-the art instrumentation (Brown et al., 2016; Hawkes et al., 2018a; Petras et al., 2017). Novel approaches for the automated analysis of high throughput data will facilitate this process (Petras et al., 2017; Rogers et al., 2019; Wolfender et al., 2019).

5.2.3 The fragmentation sensitivity continuum of DOM formulae

The four IPIMs differed in the number of isolated precursor ions ($33 - 44$), the fraction of assigned precursors (molecular formula assigned, $64 - 90\%$) and the number of product ions at CID 25 ($198 - 491$, Table A-12). All of these properties increased with mass. We calculated ion abundance-weighted averages of molecular indices based on precursor formulae. Their analysis revealed that ion mixtures at lower nominal mass differed by their higher number of double bonds (DBE), apparent higher aromaticity (AImod) and lower nominal oxidation state of carbons (NOSC, Table A-12, $p < 0.05$). Highest numbers of fragment ions were always detected at CID 25 (Figure A-3). The product ion spectra did not indicate abrupt structural changes during the increase to CID25, indicating no clear separation of groups of isomers. The four IPIMs consistently changed their molecular properties upon fragmentation: We saw a significant increase in ion abundance-weighted averages of DBE, DBE-O and AImod, whereas O/C and NOSC showed a significant decrease (Table A-12) with increasing collision energy. The H/C ratio and average numbers of C, H and O atoms per formula showed a less significant response. In general, IPIMs across all four nominal masses became apparently more aromatic and less oxidized upon fragmentation and also became more similar in terms of their molecular composition (not shown).
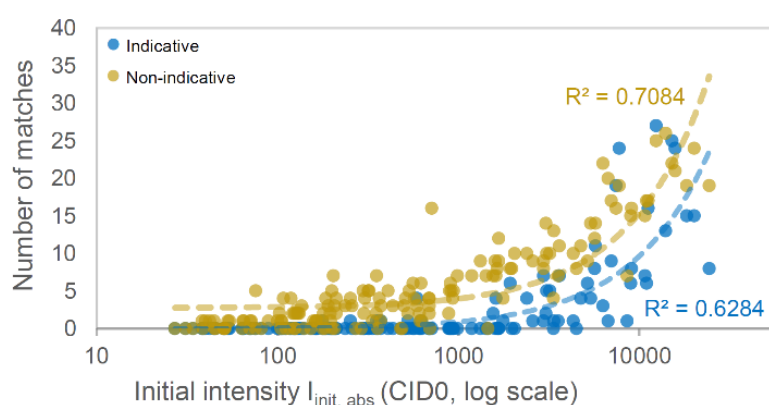
Figure 5-6. Link between precursor ion abundance and its number of MD matches upon fragmentation. Number of matches against lists of non-indicative MDs (ochre, Table A-10) and indicative MDs (blue, Table 5-1) in relation to the log initial ion abundance of the precursor. All precursors across the four IPIMs (n=159) are shown. Regression curves are linear fits (note log scale). In contrast, fragmentation sensitivity was a poor predictor of number of matches (Figure A-18).

The fragmentation sensitivity (change in intensity upon fragmentation) and number of matches to non-indicative mass differences (Table A-10) were checked on the single precursor level to assess differences between molecular formulae (*m/z* 241, Table A-13; *m/z* 301, Table A-14; *m/z* 361, Table A-15; *m/z* 417, Table A-16). We determined the fragmentation sensitivity as the relative (%) change in ion abundance based on the initial values (non-fragmented). The number of matches to the list of non-indicative MDs allowed us to relate properties such as the number of CO_2 losses to the initial ion abundance of the precursor and its molecular formula (Figure 5-4). Formulae that fragmented extensively (strong decrease in ion abundance with CID increase) had a significantly (Pearson, $p < 0.05$) lower mass defect, lower number of carbon and hydrogen atoms per formula as well as a lower DBE-O, and showed higher numbers of oxygen atoms and subsequently had a higher O/C and NOSC. Contrastingly, formulae with less oxygen and a higher mass defect were also less sensitive to increasing collision energy. This observation was in line with above described findings on the level of the nominal mass, but showed that increases in aromaticity and DBE were mostly assigned to the selective fragmentation of oxygen-containing formulae. In strong contrast, these trends did not show a significant correlation to the initial ion abundance of the precursors. However, initial ion abundance had a strong effect on the amount of matches, both to the list of non-indicative and indicative MDs (Figure 5-6). Fast-fragmenting precursors showed multiple losses of CO_2 (up to four, Figure 5-4a) or H_2O units (up to two), in line with the fast decay of formulae rich in oxygen. Contrastingly, the number of CH_2 (up to four, Figure 5-4b) losses was not correlated to the both investigated properties (fragmentation sensitivity/ match frequency).

These observations are in accordance with reported results, showing a strong dependency between oxygen content of a formula and the number of product ions (Brown et al., 2016; Capley et al., 2010; Hawkes et al., 2018a; Mawhinney et al., 2009). We assessed fragmentation sensitivity of single precursors by a second measure, denoting a 50% decrease in initial ion abundance, the half-life CID energy (derived from linear regression of ion abundance data). The average half-life CID energy of formulae with O/C ratios > 0.7 was ~ 15 while it was ~ 25 for formulae with O/C ratios < 0.2 (Figure A-20). As a result, the same continuum of fragmentation sensitivities was found at each of the four IPIMs, ranging from CID 10 – 35 under our instrumental settings (Figure A-20e). This indicates that intrinsic averaging prevailed in the property of fragmentation sensitivities in our study, similar to continua of polarity, reactivity, molecular size, or photochemical properties revealed by other studies (Arnosti et al., 2018; Brown et al., 2016; Hawkes et al., 2019; Mostovaya et al., 2017). Single formulae also showed zero or slightly positive changes in ion abundance with increasing collision energy. The respective formulae had an average O/C ratio of 0.19 and were of

low initial ion abundance (average, absolute intensity of 100) which at maximum doubled until highest applied energies. The fraction of ion abundance of these minor signals was equivalent to 0.5% of total initial ion abundance, and thus negligible. Such effects are not unexpected, as ion detection might be hampered by space-charge effects in the Orbitrap cell (Zubarev and Makarov, 2013). However, the small change in abundance of single signals documents that those effects were negligible in our analysis and affected only a group of minor signals that were insensitive to fragmentation. In contrast to IPIMs, the set of pure substances showed a remarkably narrow range of half-life energies, ranging from CID 15 – 20 (Figure A-20e). Ellagic acid was the only exception with a higher value (CID 30) caused by the high stability of two intramolecular lactone bonds.

Abundant oxygen-rich formulae contributed mostly to the tandem MS information in terms of non-indicative MDs (Figure 5-3b, and following the Van Krevelen patterns of CO_2 and CO, panels a and c in Figure 5-4, but in contrast to CH_2, panel b), in contrast to indicative MDs (Figure A-21). Precursors matching to this list were limited to a set of ~10 formulae per nominal mass (Figure 5-3b), peaking in the central area of the Van Krevelen plot (Figure A-21b) at similarly high numbers of matches per individual precursor (~30). It is thus concluded that two overlapping, but dedicated pools of formulae selectively contributed to the overall MS^2 signal, and that the usually observed fragmentation pattern is dominated by a set of highly oxidized formulae not representing the whole structural detail of the DOM mixture. Yet, these MDs are connected to the most abundant product ions. A significant pool of molecular formulae contributes to MD information that is connected to less common but potentially more indicative signals.

5.2.4 MD-enhanced Van Krevelen plots reveal structural similarities in DOM

The number of matches to the indicative MDs (Table 5-1; Figure 5-4d – f, Figure 5-5) was mainly linked to initial ion abundance (Figure 5-3b, Figure 5-6). Moreover, the matching patterns of different members of an IPIM were quite variable and surprisingly specific. Highly indicative sugar losses (glucose, rhamnose and the radical form of rhamnose, Figure 5-4e) showed an intriguing concentration in the Van Krevelen plot at relatively high values of H/C and O/C ratios, as opposed to the pattern of methyl radical loss (Figure 5-4d). Moreover, polyol losses were dominantly detected in IPIMs at *m/z* 301 and 361 (Figure 5-4e), suggesting m/z-related differences in contribution of glycoside structures. Similarly, we found that gallic acid-related MDs (either with or without retention of H_2O, and with additional loss of H_2O) concentrated in the center of the Van Krevelen plot and were more frequently associated to IPIMs at *m/z* 361 and 417 (Figure 5-4f). The

frequency pattern of the methyl radical MD paralleled the patterns of precursors indicating preferential loss of CH_2 and CO units, preferentially in the IPIMs at *m/z* 241, 301 and 361 (Figure 5-4b, f). Similar to MDs equivalent of gallic acid losses, Highest matches to MDs related to flavan-3-ol (Figure 5-5a) and flavon-3-ol structures (Figure 5-5b) were found in the central part of the Van Krevelen plot. Two molecular formulae (peaks 4 and 7 in Figure 5-3b) matched to sets of indicative MDs related to the sequential cleavage of the flavan-3-ol three-ring systems (Figure 5-5c), and peak 4 did also match to the MDs indicative of a ring cleavage of the flavon-3-ol ring systems (Figure 5-5d). In general, we observed that the Van Krevelen frequency patterns of indicative MDs differed between the four precursor ion mixtures, probably due to the fact that they each covered unique sectors and/ or lines of the VK plot (Figure 5-4) and were constrained by their mass (i.e., polyol losses not observed at *m/z* 241).

Taken together, the MD matching approach provided a more reliable base for the definition of domains in Van Krevelen space, and actually provides evidence that the domains are far from being defined; it is highly likely that formulae plotting in the "tannin" region of the plot (low H/C ratios, high O/C ratios, Figure 5-2a) are no actual tannins and that formulae in the "lignin"-like region (central, crowded area in most Van Krevelen plots; Figure 5-2) of the Van Krevelen plot link to a great variety of MDs, equivalent to e.g., gallic acid-related structures, polyol-like structures, or even indicate the potential presence of flavonoid-like ring systems, which is in accordance with the high overlap seen for different structural groups (Figure 5-2a). The application showed that polyol-expelling molecular formulae (precursors with MD matches equivalent to polyol structures) indeed showed a slight shift to the upper right (high H/C, high O/C) of the Van Krevelen plot which is the expected range of carbohydrate-like structures (Figure 5-4e). Interestingly, this area was clearly separated from precursors indicating preferential losses of CH_2, CO_2, CO and $CH_3^•$ MDs or MDs equivalent to galloyl-, flavonol- and flavanol-related structures (Figure 5-4, Figure 5-5). The important difference to usual formula-based classifications is that the MD approach used here has a tandem MS dimension that also allows to be proven by selective treatment of DOM, for example by reductive degradation, enzymatic treatment, or targeted decoration of functionalities to change properties of specific precursors (Arakawa et al., 2017; Baluha et al., 2013; Duncan et al., 2016; Linz et al., 2018; Luek et al., 2018; Nebbioso and Piccolo, 2015; Pohlabeln and Dittmar, 2015). Our approach thus opens a novel way to observe *m/z*-dependend chemical change in individual DOM samples or sets of them.

Identification of ecosystem markers by Orbitrap tandem mass spectrometry

We compared the matching patterns of different DOM precursors and standard compounds obtained at similar CID energy (Table 5-1, only DOM precursors and standard compounds with at least one match included). In general, there were always strong differences between DOM precursors, with single formulae showing high numbers of matches (Figure 5-3b, Figure 5-7a). Based on these matching profiles, we used cluster analysis to check for similarities between pure substances and DOM precursors (Figure 5-7b). The analysis showed clear clusters, grouping for example Catechin (#10) together with two precursor formulae ($C_{16}H_{14}O_6$ and $C_{17}H_{18}O_5$, which are linked by a formal exchange O vs CH_4). Similarly, $C_{15}H_{10}O_7$ was grouped with the flavon-3-ol core structures of Spiraeoside (#12) and Isoquercitin (#13) obtained by MS^2 of the intact flavonoids and subsequent MS^3 fragmentation of their flavon-3-ol cores ($C_{15}H_{10}O_7$ is also the neutral molecular formula of this core structure). Two additional molecular formulae were grouped with the intact Isoquercitin (#13) structure that showed a loss of a sugar unit (a.o.). The both precursors ($C_{13}H_{18}O_8$ and $C_{14}H_{22}O_7$) were also linked by a formal O vs. CH_4 exchange and showed a highly similar MD pattern (as opposed to other members of the series, e.g. $C_{12}H_{14}O_9$). Both formulae matched to highly indicative losses of quinic acid, rhamnose and glucose, all of which are polyol structures. Most intriguingly, this match was close to exclusive, and strongly contrasted by the formulae matching with catechin (showing no indication of a polyol loss despite their high number of MD matches).
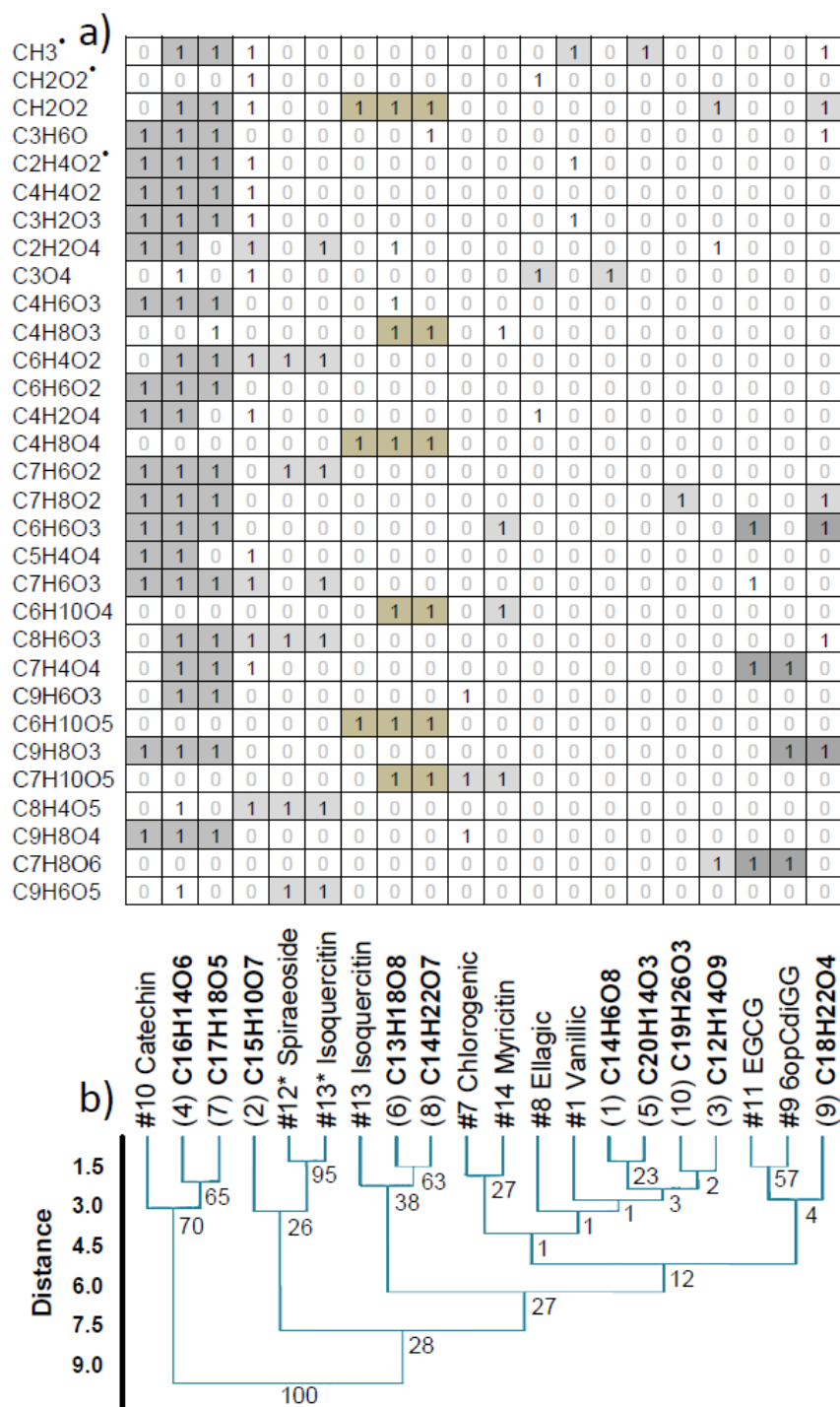
Figure 5-7. Indicative MD matching similarities among DOM precursors and standard compounds. MD matching profiles and similarity of DOM precursors of IPIM at m/z 301 and standard compounds fragmented at CID25. a) List of indicative MDs (Table 1) that showed at least one match to the DOM precursor data. Standard compounds and DOM precursors are in columns (col names, refer to b). "1" indicates the presence of a MD, "0" indicates absence; shading is only for visual guidance. Numbers in brackets refer to numbers of formulae in Figure 5-3b and #IDs refer to Figure 2-3. b) Hierarchical cluster analysis of MD matching matrix in a) (Ward's method, Euclidean distance, 999 permutations, cophenetic correlation coefficient 0.83) Percentages at nodes are indication to node stability during bootstrapping.

Other precursor matching profiles were less indicative: Precursor $C_{14}H_6O_8$ matched to only one specific MD that originally was found in ellagic acid (#10), which also shares this exact same molecular formula. However, two other MDs detected in ellagic acid ($C_4H_2O_4$ and $CH_2O_2\bullet$; Figure 5-7) did not match with the respective DOM peak. Ellagic acid did not fragment extensively at CID25, but showed six more MDs in the list of non-indicative MDs (Table 5-2), which were partly also detected in the DOM precursor but not included into the similarity analysis due to their low information content. Taking them into account led to a final match of 75% (6 out of 8 MDs), and five of those were major fragments of ellagic acid at CID25.Much similar, the four potential tannic ecosystem markers from a previous study (Roth et al., 2014) showed only matches to the ubiquitous list of MDs (Table A-10). We found no indication of gallic-acid related losses (Figure 5-4f) in these precursor's MD matching profiles. Despite the high DOC concentration of the injected sample and their high ion abundance, these peaks matched only to non-indicative MDs and not as expected, to losses of galloyl subunits. Indication for these type of losses ($C_7H_4O_4$, $C_7H_6O_5$, $C_7H_8O_6$) was instead only found in the central part of the VK plot (Figure 5-4f), which is usually assigned to lignin or CRAM-related or "lignin-like" structures (Figure 5-2a). Based on these results, the previous assignment of formulae as potential tannins, which was based solely on their position in the Van Krevelen plot(Figure 5-2a; D'Andrilli et al., 2015; Roth et al., 2014), has to be questioned. It seems rather likely that the high O/C of these molecules is caused by decoration of a central aromatic core with carboxylic acid groups. There is no question about the unique signature that the forest samples showed in the initial study (Roth et al., 2014), and similar structures are repeatedly described as markers of low-pH/high-DOC ecosystems (chapter 4; Gonsior et al., 2016; Simon et al., 2019).

It was of particular interest that pairs of formulae linked by a formal CH_4 vs. O exchange (Stenson et al., 2003; These et al., 2004) showed highly similar matching profiles in terms of their indicative MDs (Figure 5-7a, b), across all four IPIMs covered in our analysis (e.g., $C_{16}H_{14}O_6$ and $C_{17}H_{18}O_5$). Most notably, this discerned them from other members of the same series (e.g., $C_{18}H_{22}O_4$ and $C_{19}H_{26}O_3$) and not just from members of other series present at the same m/z (e.g., $C_{14}H_{22}O_7$). Initial ion abundance was not the main driver of this effect: the molecular formula $C_{12}H_{14}O_9$ showed a distinctly different matching profile as compared to $C_{13}H_{18}O_8$ and $C_{14}H_{22}O_7$, despite the fact that all three precursor ions showed high initial ion abundance (Figure 5-3b). High congruence of fragmentation patterns among sets of precursors has also been described for the MDs of H_2 and CH_4 vs. O (Capley et al., 2010; These et al., 2004), suggesting that structural differences are residing in backbone structure of similarly substituted molecular structures. The origin of the CH_4 vs. O exchange remains unknown (These et al., 2004); it may be of special interest that we observed a

minor formal insertion of oxygen for a CH_4 upon FTMS analysis in case of the both methoxyphenols (**#4** and **#5**). Full explanation of this phenomenon was not the scope of our paper and will require further systematic experiments (Baumeister et al., 2018) with regard to the identfication of this constituting DOM mass difference.

5.2.5   Link between indicative MDs, database hits and DOM chemodiversity

Although our approach isolates and fragments a mixture of ions (IPIM, isobaric precursor ion mixture) and generates so-called chimeric spectra that are not suited for single compound identification (Petras et al., 2017), complexity was greatly reduced (Figure 5-1a, b) and MD annotations added a layer of chemical information to the MS^1 data that complements the usual molecular formula information. It also allowed us to compare identities of matching MDs with structural suggestions from a natural product database, and thus opened a way to search for structural analogs. The structure suggestions from DNP agreed well with the MD predictions for all four IPIMs. For example, we found 264 structure suggestions for the catechin-like formula $C_{16}H_{14}O_6$ of IPIM at m/z 301. About 40% of these hits (n = 101) were related to flavanone structures; 43 additional structures were related to other flavonoid classes (chalcones, pterocarpanes, etc. ~16%). Another 72 (~28%) structures belonged to similar 3-ring aromatic systems (xanthones, antraquinones). In agreement with the MD matching profile and as expected from the molecular formula, no glycoside structures were suggested. Similarly, among the 174 hits for the formula $C_{17}H_{18}O_5$, about 26% (n=44) were related to flavan structures, agreeing well with the similarity to catechin in our study. The majority of other hits were related to aromatic systems with three rings (n=49, 28%; 3-ring systems as xanthones or antraquinones), or two-ring systems derived from lignan, stilbene, benzophenone or coumarin base units (n=76, 44%). No glycosides were suggested. Of the 129 hits that were obtained for molecular formula $C_{15}H_{10}O_7$, 56% (n=72) were structures containing a flavone core, in agreement with our matching profile result. Twenty structures were flavonoid-related (~15%, e.g., aurones, pterocarpans) and 36 structures (28%) belonged to related three-ring systems (mostly, anthraquinones). Finally, in the case of the both formulae matching to polyol MDs, less numbers of structures were suggested in DNP ($C_{13}H_{18}O_8$, 16; $C_{14}H_{22}O_7$, 7). DNP suggestions agreed well with matching predictions in the first case (all suggestions were mono-glycosides of an aromatic ring), but less well in the latter case (aliphatic chain structures, though in one case linked by a central polyol structure, but more often linked by intramolecular ether, ester or lactone functionalities). In case of formula $C_{18}H_{22}O_4$, which matched to MDs equivalent to caffeoyl, coumaryl and phloroglucinol moieties (Figure 5-5b), 28 of the 74 suggestions were related to

lignans or similarly linked aromatic structures: many of those structures contained C_1C_6 or C_3C_6 structures (such as vanillyl or coumaryl moieties) that could give rise to the observed MDs.

Our approach does not lead to the identification of an unknown in the classical sense. It rather provides a novel way to describe and analyze mixtures of precursor ions (IPIMs) that themselves represent isomeric mixtures (Brown et al., 2016). This does not exclude the possibility to identify single compounds under specific, targeted conditions. However, these conditions will not only require tailored analytical and (bio-)informatics tools but also detectable amounts of a "pure" isomer; a circumstance that may be less probable as expected in the natural environment (Brown et al., 2016; Hawkes et al., 2018a; Petras et al., 2017; Zark et al., 2017; Zark and Dittmar, 2018). The analysis of MDs in DOM will thus benefit by process-oriented experimentation and/ or strong environmental gradients that may reveal molecular responses properly (Boye et al., 2017; Mentges et al., 2017; Moritz et al., 2017; Petras et al., 2017; Zhang et al., 2014). For example, the molecular formula $C_{16}H_{14}O_6$ (m/z 301.0717) was one of the most intense ions at that nominal mass and matched to many indicative MDs (Figure 5-3b, Figure 5-7a, b). It showed a high degree of overlap with the catechin MD fingerprint which were both measured at CID 25 (Figure 5-7a, b). Structural suggestions from DNP pointed into the same direction, with about 56% of hits (n=144) related to flavonoid scaffolds, and another 28% of additional hits covering related three-ring scaffolds not included by our study.

We found a strong link between the number of matches to the indicative MDs and the number of structural suggestions by DNP (n, precursors = 48, p = 0.004 **, p = 0.117 for all matches incl. non-indicative MDs), which became more significant when both variables were rank-transformed to account for extreme cases (***, p < 0.000; p = 0.03 for all matches incl. non-indicative MDs). At this point it is not clear whether the increased matching frequency in the center of the Van Krevelen plot – which is also linked to most intense and ubiquitous precursor ions (Hertkorn et al., 2007; Kellerman et al., 2014; Roth et al., 2014) – is a sign of a high underlying isomeric diversity or if it is just due to better S/N of these precursors and their product ions. Our MD analysis considered any MD between all precursor ions and all product ions; from a simple stochastic viewpoint, every precursor had the same chance to match to the both lists of MDs (Table 5-1, Table A-10). The observed patterns differed markedly from such a random result; and the high robustness of the analysis (Figure A-17) supports that a biogeochemical signal is the cause of the emerging pattern. It was also intriguing that those molecular formulae with high numbers of matches to the list of indicative MDs showed a strong tendency to yield higher number of hits in DNP. The low coverage

of DOM formulae in databases of any kind has been reported frequently, and generally leads to total assignments of less than 5% of formulae to known structural suggestions and/ or biochemical pathways (Brown et al., 2016; Petras et al., 2017; Zhang et al., 2014). In line with these results, we saw strong differences among queried formulae in DNP; however, the number of molecular formulae in our subset showing at least one hit was quite high (41 out of 48, i.e., 85%). Formulae below mass defects of 0.05 (O-rich, H-deficient) had one hit (n=13, median) while formulae above had 33 (n=35, median), similar to previous results stating that especially oxygen-rich formulae were under-represented in SciFinder (Brown et al., 2016). The number of matched indicative MDs that we assessed in this study could be interpreted as a first, very rough measure to account for underlying isomeric and molecular complexity of a precursor because we excluded all those MDs that are deemed non-indicative (i.e., CO_2, etc.). Our observations would then agree with theoretical considerations on the probability distribution of structural diversity in two-dimensional Van Krevelen space (C-H-O space; Hertkorn et al., 2008, 2007; Petras et al., 2017; Reemtsma, 2010). We particularly want to acknowledge at this point that there may be a high number of unknown structural scaffolds which are not yet part of natural product databases, such as DNP (Chassagne et al., 2019). Similarly, not every chemically meaningful MD was included into our analysis, a task that would require further exploitation of available tandem MS databases and DOM data (Petras et al., 2017).

5.3   Conclusion and outlook

This study represents a first step to improve our understanding of the complete chimeric tandem MS signal emerging from ultra-complex mixtures which resist any classical means of separation, isolation, and identification. In line with recent experimental and theoretical considerations, we found an imprint of remaining ecosystem (source) signals in soil dissolved organic matter (DOM) that continuously vanishes by transformation during ecosystem passage and along aquatic gradients. Tandem MS-enhanced 3D-Van Krevelen plots showed a separation of precursor formulae contributing to indicative losses (e.g., methyl radical, polyol, or gallate losses). This analysis indicated blurry domains that were inconsistent with previous a-priori classifications (see e.g., chapters 3 and 4), which have been criticized in the past. Our mass difference (MD) matching approach yielded consistent and robust matching results that suggested a dominance of neutral losses from the dilute precursor mixtures and allowed for deconvolution of precursor "MD match spectra" from complex isolated precursor ion mixtures (IPIMs). However, these assumptions have to be further tested with mixtures or standard compounds. Tandem MS databases and novel analysis

tools will help in achieving these aims. Comparative metabolomics of chimeric tandem MS fingerprints selected by hypothesis-driven experimentation offered new insights into ecosystem-specific DOM structures. It thus emerges as a workflow that can complement the existing broadband MS[1] analysis protocols by adding meaningful structural information. In line with predictions from the matching approach, we found consistent structural suggestions in a large natural product database. Our results support the theoretical consideration that formulae in the central part of the Van Krevelen plot do show highest realized isomeric complexity due to the high degree of freedom in structural space, as dictated by stoichiometry and valence, and thus suggest that our approach could help to reveal "limits" of chemodiversity per molecular formula. Altogether, the proposed MD matching approach opens up new avenues to access DI-ESI-Orbitrap MS/MS data (and potentially, also LC-MS/MS data) of complex mixtures and the information encrypted therein.

6   Synthesis

This thesis provides novel insight on the use of the Orbitrap mass spectrometer at its highest commercially available resolving powers (200k – 500k) for the detailed analysis of ecosystem-specific compounds in complex mixtures of terrestrial dissolved organic matter (DOM). The findings of this thesis contribute to a better molecular understanding of global biogeochemical cycles with a focus on the recycling and transport of DOM in soils and how this links to aquatic systems. We showed 1) that a simple "one sample/ whole spectrum" tuning approach improves the between-lab reproducibility of FTMS instruments and that the Orbitrap emerges as a suitable alternative to state-of-the-art FT-ICR-MS instruments (**chapter 3**); 2) that two contrasting whitesand (podzolic) sites in the Rio Negro basin differ strongly in their DOM fingerprint and that highly oxidized aromatic markers of dry upland Podzols show strong overlap with sets of previously described Rio Negro DOM markers and putative tannin markers of temperate conifer forests (**chapter 4**); and 3) that the direct injection tandem MS analysis of selected putative tannin markers in soil water DOM from acidic forest sites opens a new avenue to assess and compare molecular ecosystem imprints (**chapter 5**). These findings provide new impetus in terms of instrumental and experimental approaches to improve our understanding of DOM dynamics at different scales (soils, catchments, river networks). Our results thus contribute to the assessment of unknown markers in fast-responding but poorly understood terrestrial DOM mixtures, and the broader implications of these findings will be discussed in the following.

6.1   The Orbitrap analyzer reveals ecosystem markers in terrestrial DOM

FTMS technology has greatly improved our understanding of global DOM dynamics (Dittmar and Stubbins, 2014; Minor et al., 2014; Roth et al., 2015), yet we stand at a point where novel breakthroughs are required to reach overarching interpretations from the large amounts of data that FTMS methods generate. Methods that enable the observation of DOM properties beneath the level of the molecular formula, i.e., that enable us to discern isomers, are urgently needed to do so (Hertkorn et al., 2008; Petras et al., 2017).   The low accessibility of FT-ICR MS instruments dedicated for the non-targeted broadband analysis of DOM hinders progress of the field (Hawkes et al., 2016). Orbitrap analyzers have emerged as alternatives over the past decade (Denisov et al., 2012; Hawkes et al., 2016; Zhurov et al., 2013). This work has demonstrated the potential of the powerful high-field Orbitrap analyzer to investigate various

forms of terrestrial DOM and ecosystem markers therein. Although it is evident that resolving powers are still too low to differentiate specific mass splits related to nitrogen, sulfur, phosphorus, or rare isotopes (^{13}C, ^{15}N, etc.) and especially their manifold combinations, it is reasonable to assume that these molecular formulae relate to rare and specific metabolites that do not represent the majority of oxidized, acidic and soluble DOM molecules found in natural waters (Hertkorn et al., 2008; Minor et al., 2014). While larger lab comparisons have been launched in the field of clinical metabolomics (Martin et al., 2015), such initiatives have been missing in the field of DOM metabolomics (Hawkes et al., in prep.). This depicts that the instrument response – upon ionization and injection of an ultra-complex sample such as DOM – is not understood well yet (Cao et al., 2016; Novotny et al., 2014). Unexpectedly, our simple "one-sample/ whole-spectrum" Orbitrap tuning approach against a reference 15T FT-ICR MS instrument resulted in > 90% signal overlap by commonly detected species for a broad variety of samples. This indicates that both instrument types are essentially detecting the same complex biogeochemical signal and are able to draw a similar degree of information from it, which subsequently leads to a reproducible biogeochemical interpretation. This is a precondition for future comparison studies among labs to reach overarching conclusions.

The higher reproducibility of the Orbitrap which was also noted in other studies (Hawkes et al., 2016) makes it especially suitable for marker identification. Low measurement variation is key to differentiate responsive DOM signals by comparative metabolomics, e.g., when focusing on differences between ecosystems or experimental treatments (Kuhlisch and Pohnert, 2015). However, detection of novel ecosystem or process markers also needs proper knowledge of which signals contain no information, besides measurement variation (instrumental noise), i.e., "non-informative" process and source variation needs to be known. This aspect, however, is highly scale-dependent, especially when one deals with the "ecosystem" concept which is applied from the micro- to the macroscale (Cameron and Earley, 2015), and especially when looking at ubiquitously present DOM that also interlinks "systems". Much of the ambiguity between sets of indicative FTMS process and source markers (see section 1.4.1) may thus relate to different scales covered and limited representativity of lab and field experiments. Openly-available molecular-level information on DOM and the related metadata – e.g., experimental setup, scope of the study, details on sampling and measurements, and instrument type – are still scarce (Petras et al., 2017) and hinder progress. Future efforts should deliver such data to reach overarching, instrument- and lab-independent sets of informative markers for a given process (DOM modification) or source (DOM type) (Petras et al., 2017; Wang et al., 2016). The

Orbitrap analyzer will allow a wider research community to contribute to this process (Hawkes et al., 2016).

A major drawback of all FTMS methods is their inability to quantify the amount of ionized DOM and effects of different ionization efficiencies. This is especially critical in regard of studies showing that ESI-FTMS detection misses unknown parts of the DOM pool, which may bias our interpretations (Hawkes et al., 2019; Rodgers et al., 2019). The understanding of the observed DOM signal in state-of-the-art FTMS analyzers will remain qualitative unless novel ways of signal monitoring in the front end instrument are developed. Preparative mass spectrometry could open possibilities to recover the ionized material in a non-destructive manner and to simply compare concentrations of carbon (nitrogen, sulfur, etc.) before and after ionization (Gologan et al., 2004; Johnson et al., 2016). This strategy could also be of value to decide on the "correct" ionization mode and mass window to measure DOM (i.e., which window or ionization mode reflects best ionization/ measurement conditions of the sample). For example, it is evident that sample types (e.g., bog and aquifer) differ in their spectrum envelopes, i.e., ion abundance patterns, but it has also been shown repeatedly that the spectrum envelop is highly impacted by the accumulation time prior to FTMS measurement (Cao et al., 2016; Hawkes et al., 2016). Such effects are poorly understood and rarely addressed by the FTMS community, and thus constitute major challenges for the future application of these powerful instruments (Nielsen et al., 2018).

6.2   A molecular picture of DOM ecosystem imprints and links

The combination of datasets from different instruments can benefit the search for ecosystem-specific DOM markers, their sources and the development of new hypotheses on their fate at different scales. For example, SOM and DOM dynamics are strongly coupled in terrestrial systems (Figure 1-2; Figure 1-3; Figure 6-1) but the molecular dimension of this "coupling" and its mechanisms are poorly understood (Malik et al., 2016; Roth et al., accepted). Two groups of potential markers were identified in this thesis: markers of "fast-cycling" high-DOC environments such as bogs, blackwaters, and topsoils, and those of "slow cycling" low-DOC environments, such as deep soils and aquifers. The differentiation into fast- and slow-cycling environments emerges from the strong differences in DOM fingerprints we observe between these both types of systems at the scale of the global water cycle (chapter 3).
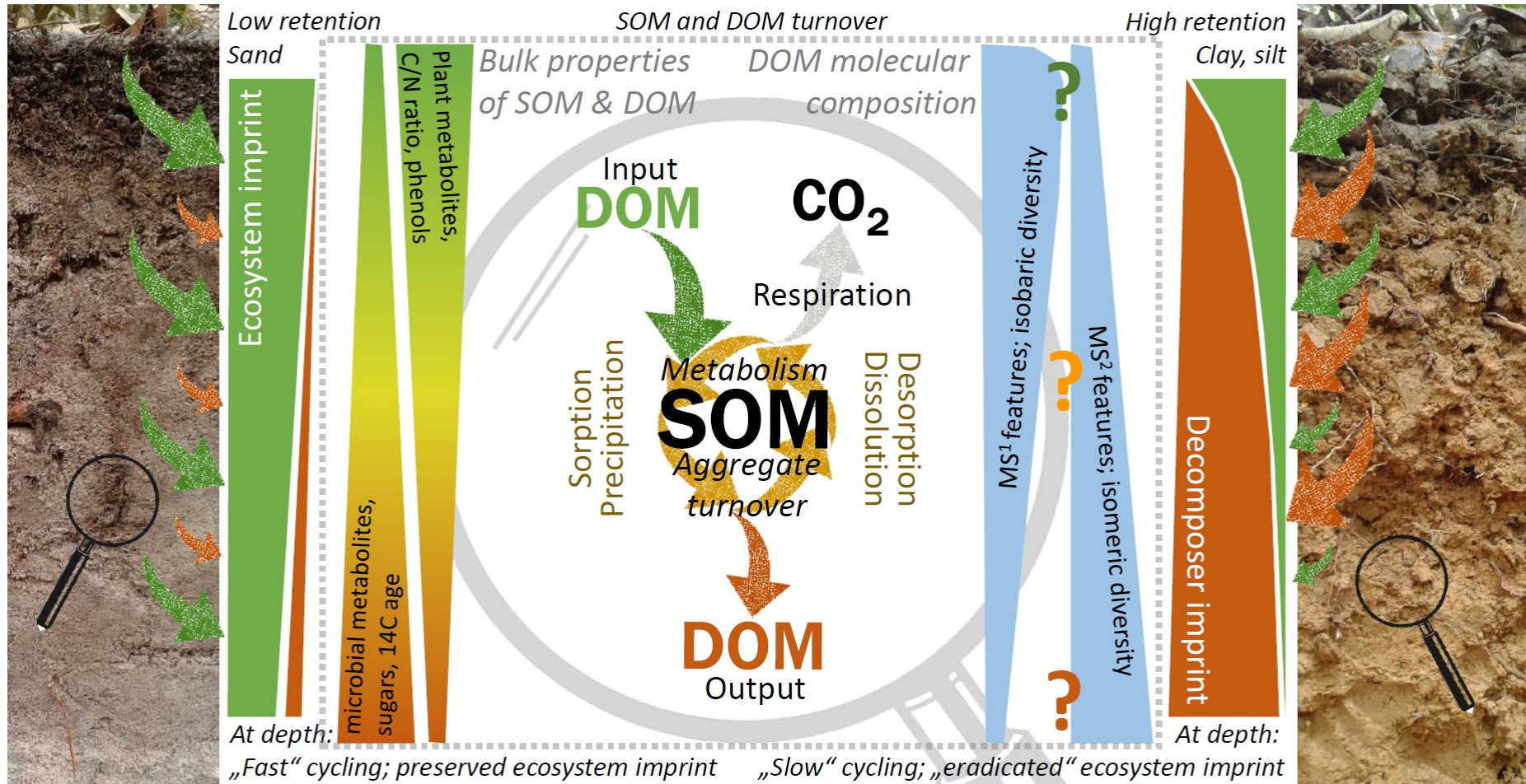
Figure 6-1. A molecular picture of ecosystem imprints in DOM (adapted from Kaiser & Kalbitz 2012; grey central box). Soil profiles at left (sandy Podzol, Campina site at ATTO, Brazil) and right (clayey Oxisol, plateau north of ATTO) show the influence of soil texture on DOM preservation or "eradication" of topsoil ecosystem imprints. Both profiles have an approx. depth of ~ 30 cm. Columns only show trends. See main text for explanation.

The role of soil texture in determining the intensity of recycling on much smaller scales, such as the Critical Zone, became apparent as well: Sandy soils (Figure 6-1, left soil profile) allow quick transport of DOM from fast-cycling environments, e.g., topsoils, to more "isolated" environments beneath, e.g., shallow aquifers, or deeper soil, and thus favor a delay of the recycling process and thus prolong the preservation (or, persistence) and restricted "eradication" of topsoil ecosystem signals (see further below; Küsel et al., 2016). This shows, one the one hand, that FTMS methodology reveals ecosystem imprints at both global and local scales, but on the other hand it also shows that the resulting molecular patterns may be ambiguous and need to be studied at much higher detail.

Aromatic and phenolic DOM molecules emerged as markers of input-dominated systems (chapter 3), in line with previous findings (Figure 6-1, "ecosystem imprint"), however, a special category of them has been repeatedly described as markers of low-pH/ high-DOC environments such as forest soils, bogs, and blackwater rivers (chapter 4; Gonsior et al., 2016; Roth et al., 2015, 2014; Simon et al., 2019). Moreover, we have shown clear evidence for a potential molecular link between highly oxidized aromatic formulae of relatively high molecular weight – theoretically classifiable as "tannins", in line with Roth et al. (2014) – between whitesand headwater ecosystems and the downstream river, the Rio Negro. This whitesand/ blackwater link has been presumed (Bardy et al., 2011; Ertel et al., 1986; Leenheer, 1980) but never shown in such high detail. Our findings demonstrate that specific soil environments are the dominant source of highly decomposed Rio Negro DOM markers. This, however, represents only a first step to elucidate their identity, origin and mode of formation, which may relate to organic matter processing at the forest floor and/ or active podsolization processes (Do Nascimento et al., 2008; Lee et al., 2018). Low pH environments favor the dominance of fungi (Rousk et al., 2010), and fungi have been discussed as agents of SOM and DOM modification, both by enzyme-related hydroxyl radical formation (Waggoner et al., 2015; Waggoner and Hatcher, 2017) or by ensembles of ligninolytic enzymes (Zavarzina et al., 2018, 2011) that can not only lead to color loss and degradation but also synthesis of higher-molecular weight structures (Waggoner and Hatcher, 2017; Zavarzina et al., 2018). The release of specific aromatic compounds could also relate to ongoing dissolution processes of iron oxide-dominated plateau soils (Bardy et al., 2011; Do Nascimento et al., 2008; Fritsch et al., 2011, 2009), in line with studies presenting evidence for specific scavenging/ release of aromatic molecules by iron oxides in other settings (Coward et al., 2018; Linkhorst et al., 2016; Riedel et al., 2013). This would imply that the molecular formulae connected to this specific molecular

imprint could be useful as markers of the continuing landscape degradation and potential loss of highly diverse *terra firme* forests (Coward et al., 2018).

Besides their unknown origin, our results imply that specific DOM markers may allow to discern the influence of dedicated whitesand systems *at the molecular level* in the case of the Rio Negro basin, as was suggested earlier (Bardy et al., 2011). This may open new ways to partition DOC fluxes of plateau sites, riparian corridors, and whitesand areas on the watershed scale (Alvarez-Cobelas et al., 2012; Hernes et al., 2017; Simon et al., 2019; Webb et al., 2018). An observation that links to previous findings (Roth et al., 2015) was the difference of DOM signatures between whitesand sites despite relatively similar pH; our findings imply that higher-order controls such as microclimate that eventually control, e.g., amounts of precipitation, frequency of dryness, water retention, and oxygen depletion in stagnating subsurface water bodies, have a stronger impact on DOM signatures than the pH. It rather seems that the amount of DOC in such ecosystems is the main determinant of acidity due to the lack of other active pH buffers (Markewitz et al., 2001). Anoxic conditions in poorly drained riparian systems likely explain their distinct imprint of saturated N- and S-containing formulae. Anoxic conditions have been described before in blackwaters (Goulding et al., 1988; Sioli, 1954) and require further attention, especially looking at the larger differences among this and other Rio Negro studies (Gonsior et al., 2016; Simon et al., 2019) in terms of N and S-containing formulae, which may relate more directly to cell metabolism (Kujawinski, 2011).

The second group of emerging molecular markers was indicative of slow-cycling low-DOC environments (Figure 6-1, "decomposer imprint") and characterized by the near-complete absence of aromatic and phenolic constituents and pronounced abundance of saturated aliphatic structures, such as lipids, peptides or decomposed sterols and terpenoids, showing a strong shift towards higher molecular weights (> 400 Da). The dominance of this signature is intriguing because it aligns well with the presumed microbial imprint in environments characterized by low DOC concentration such as deeper soil that subsequently also feeds into aquifers (Kaiser and Kalbitz, 2012; Kästner and Miltner, 2018; Woolf and Lehmann, 2019). In this picture, subsoil emerges as a reactor were the topsoil ecosystem DOM imprint (topsoil, dominance of plant-related phenolics, Figure 6-1, green arrows) is "eradicated", and gives rise to a decomposer imprint that becomes ever more dominant with depth (red arrows in the figure; Roth et al., accepted). Fungal DOM decomposition, as explained above, may be one possible mechanistic explanation for such a mass shift (Waggoner and Hatcher, 2017; Zavarzina et al.,

2018); ionization effects that confound the depth gradient in direct-injection broadband FTMS studies may be another. Further work is needed to rule out such effects, as noted in section 6.1 (Hawkes et al., 2019; Rodgers et al., 2019). The "eradication" process depends on the amount of precipitation entering soil and the soil texture (sandy vs. clayey, Figure 6-1) because it likely determines the efficiency of the recycling process, e.g. amounts of DOC leached, contact times between soil and water, and decomposer community composition (Krueger et al., 2018; Lee et al., 2018; Peyton Smith et al., 2014; Remington et al., 2007; Schlüter et al., 2019). In the Rio Negro basin, the combination of high precipitation, good drainage of sandy soils, and high production of litter leads to the unique setting were decomposition may be delayed to a degree that large amounts of specific ecosystem DOM escape to the aquatic continuum. In rather poorly drained soils, ecosystem signatures may vanish quickly below topsoil (Klotzbücher et al., 2016; Roth et al., accepted). However, further field experiments are needed to disentangle such molecular level ecosystem effects and depth trends in the tropics.

We unexpectedly found very young DOC in both whitesand systems that paralleled the escape of relatively "fresh" topsoil DOM to the aquatic continuum. Although presumed by most authors, our study is the first direct and simultaneous observation of DOM composition and DOC radiocarbon age (Bardy et al., 2011; Ertel et al., 1986; Leenheer, 1980) that reveals this link. Previous findings from Amazonian rivers revealed apparent fast cycling (< 5 years) of carbon between C fixation, DOC release, and DOC decomposition (Mayorga et al., 2005). Our study revealed that this young DOM shows a strong imprint of aromatic and phenolic constituents in case of the Rio Negro. Such molecules are prone to microbial and photochemical degradation (Creed et al., 2015; Dittmar and Stubbins, 2014; Rossel et al., 2013; Stubbins et al., 2010; Ward et al., 2013). The preferential loss in young and aromatic DOM leads to a shift towards higher age and more aliphatic character of DOM along the aquatic continuum (Kellerman et al., 2018), and thus overlaps with the group of FTMS-based markers of slow-cycling low DOC environments described above. The "eradication" of ecosystem imprints in soil and aquatic systems has also been linked to a decline in diversity-related measures, leading to more similar forms of DOM (Mentges et al., 2017; Roth et al., accepted). A similar age and similarity effect has been observed upon coagulation in water processing (Raeke et al., 2017). Taken together, our data and the literature suggest that ^{14}C-old and recycled material from terrestrial environments could contribute significantly to the "hidden" pool of aged carbon in aquatic DOM (Dean et al., 2019; Follett et al., 2014) that may be, at least initially, hidden beneath large amounts of more diverse and labile forms of young

ecosystem-specific DOM. However, our study falls short in reporting ^{14}C ages for the plateau (low-DOC) environments that presumably export low but constant amounts of such aged DOC. The above hypothesis would thus require further testing. For example, a simple DOC balance could be based on the distribution of soils and river types across Amazonia (McClain et al., 1997; Remington et al., 2007) to test whether the DOC flux from so called "clearwater" streams draining from low-elevation and low-DOC environments could sustain the riverine export and amount of old DOC in marine systems (Dittmar and Stubbins, 2014). At this point, it appears questionable whether topsoil ecosystem imprints find their way to the ocean, in light of a potential signal "eradication" in soils (Klotzbücher et al., 2016) and aquatic systems (Creed et al., 2015; Hutchins et al., 2017); the Rio Negro watershed is clearly an exceptional case, as detailed above. This explains why previous attempts to quantify the terrestrial organic matter flux by classical terrestrial plant-biomarkers such as lignin failed (Medeiros et al., 2016; Riedel et al., 2016). In contrast, the terrestrial imprint may be more of a decomposer imprint originating in deeper soil and aquifers, and it is unknown whether this imprint is more likely to persist in aquatic systems than the young, aromatic signature (Kellerman et al., 2015; Marín-Spiotta et al., 2014; Schmidt et al., 2011; Ward et al., 2017).

6.3   New avenues in identifiying ecosystem markers

A main part of this thesis was concerned with the identification of previously described conifer forest markers with a potential tannic origin (Roth et al., 2014) by direct-injection Orbitrap tandem MS (MS²). The work presented in this thesis suggests that they can be understood more generally as markers of fast-cycling high-DOC environments (chapters 3 and 4; see section 6.2). Our attempt to confirm the tannic identity of four selected markers failed: There was no sign of indicative losses of tannins, e.g., to potential gallic acid groups or sugar units. This finding, although restricted to a small number of exemplary markers, shows that the typical Van Krevelen plot-based annotation of structures can be misleading (Davies et al., 2015; Reemtsma, 2010). Despite these limitations, our initial approach layed a foundation to exploit chimeric mass spectra by, e.g. deconvolution, in the future, and the implications of these findings will be discussed in the following.

Through the molecular-level study of DOM by FTMS instruments for the last ~twenty years, the research community now faces a situation where it appears that DOM characteristics and degradation patterns may be universal and converging (Figure 6-1, blue columns, "DOM

molecular composition"): The decomposition of organic matter leads to DOM products that are less diverse in terms of MS^1 signals, less aromatic and more aliphatic, and more aged (Kellerman et al., 2018; Mentges et al., 2017). At the same time, this process leads to a seemingly rapid diversification in the isomer dimension (Zark et al., 2017; Zark and Dittmar, 2018). This means, however, that the event horizon of MS^1 mass spectrometry is reached: DOM becomes indistinguishable with ongoing degradation, and fine-scale ecosystem imprints vanish (Hertkorn et al., 2008). It is not fully clear whether these findings also apply to the soil environment (Figure 6-1), but recent findings suggest this (Roth et al., accepted). Ecosystem imprints that are found on the global scale, i.e., between input-driven, fast-cycling and recycling-driven, slow-cycling terrestrial systems, on a global (chapter 3) and regional (chapter 4) scale are overlapping to a high degree when only considering the MS^1 information. Likewise, ecosystem imprints between temperate and tropical sites do show strong overlap (similar differences between sandy and clayey sites, chapter 4 vs. Roth et al., accepted; and similarities among low-ph/ high-DOC environments, chapter 4 vs. Roth et al., 2014). These similarities across scales and climate zones suggest convergent paths of DOM decomposition that are not discernable by MS^1. We showed that MS^2 experiments are a potential tool to study the decomposition process in higher detail when exploited fully, i.e., when rare mass differences are exploited: The diversification process in the isomer dimension obviously results in the dominance of common losses that are common and thus non-informative, however, the neglection of the diverse and specific but rare (i.e., less abundant) signals will miss large parts of the potentially essential MS^2 information.

Our approach embraces the unavoidable ubiquity of inseparable signals in DOM: Clean fragmentation spectra from single isomers are the exception, and usually lead to total annotation rates < 5% (Brown et al., 2016; Petras et al., 2017; Zhang et al., 2014). In contrast, fragmentation spectra originating from mixtures of precursor ions (isobars and isomers) – chimeric mass spectra – are the rule. We thus introduced the MS^2 concept of the isobaric precursor ion mixture (IPIM) which can be analyzed analogous to the initial MS^1 information. This approach has been inspired by previous studies (Osterholz et al., 2015; Smirnov et al., 2019; Witt et al., 2009). Our approach complements targeted LC-MS based studies that aim to reveal "clean" fragmentation patterns (Boiteau et al., 2019; Kügler et al., 2019) and links to the concept of MS^2 motifs (Rogers et al., 2019). Although this approach will not lead to the direct identification of single metabolites, it can still assist in well-planned experimental setups or

environmental gradients to reveal, step by step, ever narrower subsets of structural motifs, their likely origin, and their subsequent targeted isolation.

Our finding of diverse ecosystem imprints, i.e., indicative motifs, in initially decomposed soil DOM is in line with the theory of universal DOM degradation detailed above. Indicative losses like the methyl radical (lignin-related), polyol units (cellulose-related), gallate-related losses (tannin-related), or sequences of losses related to flavanols and flavonols (plant metabolites) all support the idea that these known classes of biomolecular structures are transformed and included into a supramolecular DOM association (Hawkes et al., 2019; Kögel-Knabner, 2017, 2002; Wells and Stretz, 2019). The thesis has substantiated the observation that this ecosystem imprint vanishes during reactive transport (Roth et al., accepted) and mirrors transformation in SOM as detailed above (Kaiser and Kalbitz, 2012; Kallenbach et al., 2016). Our holistic MS² approach is especially suited to follow this initial decomposition process because it opens an avenue to study changes in *isomeric* complexity along the way. It has also become apparent that the clear differences among ecosystem imprints will be "eradicated" under surface conditions, unless DOM is preserved through external factors such as anoxia or fast transit from soils to the aquatic systems, as seen in our study and in the literature (see above). As detailed above, the Rio Negro basin emerges as a natural laboratory where such decomposition processes are delayed on large spatial scales and thus may be especially suited for further studies.

The forest markers described by Roth et al. (2014) likely persist only on short timescales when conditions are less preservative. It is thus important to evaluate the persistence of the identified markers to make them applicable for source or process identification. It appears especially critical to also identify markers of slow-cycling low-DOC environments, even though their "ecosystem" imprint may then be restricted to study DOM processing on larger scales (i.e., > river basin or watershed scale; Hansman et al., 2015; Hawkes et al., 2019; Kellerman et al., 2018). The inclusion of low-ion abundance signals of tandem MS data has the potential to reveal similarly large amounts of ecosystem information as the analyses of broadband MS[1] data that were initiated 22 years ago (Fievre et al., 1997). The supposedly high isomeric complexity of most abundant DOM signals due to convergent, multiple decomposition steps (Zark and Dittmar, 2018) needs to be tested, and our approach showed promising results in this regard. Our findings support the hypothesis that these ubiquitously found molecular formulae are characterized by maximum isomeric complexity and thus chemodiversity (Hertkorn et al.,

2007; Kellerman et al., 2014; Roth et al., 2014). This is an indication that the Orbitrap and more specifically MS^2 experiments are suitable to study these highly complex isomeric mixtures at the moment of their formation in "eradication" zones in soils and aquatic systems. Extending the set of indicative MDs through tandem MS libraries such as GNPS could further expand the diversity estimate (Hartmann et al., 2017; Rogers et al., 2019). If complex natural mixtures such as DOM represent a subspace of "realizable" chemodiversity, they may be of use to indicate gaps in natural product databases (Chassagne et al., 2019).

Our approach complements targeted LC-MS^n approaches that are better suited for "true" compound identification. Most importantly, the capability to deconvolute mixed MS^2 signals needs to be benchmarked against LC-MS^n data of known precursor mixtures, to avoid misinterpretation. Lastly, the work presented herein falls short in applying the developed approaches in concert (MS^1 and MS^2 profiling) to a set of replicated ecosystem samples (Figure 6-1). This would be needed to prove two things: First, it would answer the question whether ecosystem imprints, e.g., as determined by forest type (broad-leaved vs. conifer; beech vs. pine) prevail at different scales (site/ region) and thus across climate and soil gradients. Secondly, it would answer the question whether initial (topsoil) ecosystem imprints are better revealed by either MS^1 or MS^2 features, or if both levels of information complement each other (Figure 6-1).

Summary

The production and export of terrestrial DOM is intimately related to processes in soils and the critical zone, and society relies on the sustained functioning of these systems (Küsel et al., 2016). Understanding mechanisms of SOM turnover and DOM production are thus crucial for proper management of ecosystems. Advancements in analytical technology – namely in soft ionization techniques coupled to ultrahigh resolution Fourier transform mass spectrometers – have revolutionized our view of ubiquitous and ultra-complex organic mixtures and have made the analysis routine in biogeochemistry (Minor et al., 2014; Smith et al., 2018; Zark and Dittmar, 2018). Knowledge about terrestrial DOM remains scarce, compared to advances in the fields of aquatic and marine sciences. Although terrestrial DOM is now seen as a valuable indicator and integrator of fast ecosystem responses (Roth et al., 2015, 2014), our understanding of molecular ecosystem imprints and potential new source or process markers is still limited. Therefore, this thesis investigates the following points:

1) Evaluate the suitability of the Orbitrap analyzer as a cheaper alternative to state-of-the-art FT-ICR MS instruments for the analysis of terrestrial DOM and development of an approach to improve the reproducibility among instrument signals (**chapter 3**).

2) Apply the Orbitrap method in the context of the Rio Negro basin, i.e. extract sets of markers based on sample sets from DOM-rich whitesand areas linked to blackwater production, and further test their potential overlap with basin-scale tannin-like molecular markers reported by independent studies (**chapter 4**).

3) Develop a new approach to study ecosystem imprints in DOM by means of comparative tandem MS analyses of known polyphenolic structures, and apply it to identify putative tannic markers within forest soil water DOM (**chapter 5**).

In **chapter 3**, we developed an approach to improve reproducibility in the FTMS response of complex DOM samples, and proved that the Orbitrap is suitable for the analysis of DOM along terrestrial-aquatic gradients. We tuned the Orbitrap with a simple "one sample/ whole spectrum" tuning approach that effectively minimized differences in ion abundance patterns of DOM data. We then compared the Orbitrap method to an established 15 Tesla FT-ICR MS on a diverse suite of 17 mainly terrestrial DOM samples regarding 1) ion abundance patterns, 2) differential effects of DOM type on information loss, and 3) derived biogeochemical information. We show that the Orbitrap provides similar information as FT-ICR MS, especially

for compound masses below 400 m/z, and is mainly limited by its actual resolving power rather than its sensitivity. Ecosystems that are dominated by inputs of plant-derived material, like DOM from soil, bog, lake and rivers, showed remarkably low average mass to charge ratios, making them also suitable for Orbitrap measurements. The additional information gained from FT-ICR MS was highest in heteroatom-rich (N, S, P) samples from systems dominated by internal cycling, like DOM from groundwater and the deep sea. Here FT-ICR MS detected 37% more molecular formulae and 11% higher ion abundance. However, the overall information content, which was analyzed by multivariate statistical methods, was comparable for both data sets. Mass spectra-derived biogeochemical trends, for example the decrease of DOM aromaticity during the passage through terrestrial environments, were retrieved by both instruments. We demonstrate the growing potential of the Orbitrap as an alternative FTMS analyzer in the context of challenging analyses of DOM complexity, origin and fate. Further work is needed to constrain which fraction of DOM is ionized, to allow improved quantitative estimates based on FTMS data.

In **chapter 4**, we found evidence for the hypothesis that whitesand areas of the Rio Negro basin leave distinct molecular imprints within exported DOM that matches distinct highly oxidized aromatic (i.e., potentially tannic) DOM markers found at the basin scale. We compared site-specific molecular fingerprints from groundwater transects in two contrasting micro-catchments draining into the lower Rio Negro with published datasets on Rio Negro DOM. Water isotopes (^{18}O, ^{2}H) and carbon isotopes of DOC (^{13}C, ^{14}C) were used to constrain water sources and DOC origin and age. DOM fingerprints were analyzed by Orbitrap mass spectrometry and compared by multivariate statistics. We show that a dry precipitation-driven whitesand ecosystem (Campina forest) contributes preferentially to a major molecular DOM imprint of potential tannic Rio Negro-specific structures, despite an estimated twofold lower discharge (but 2-3 times higher DOC export) of the ecosystem as compared to a nearby wet and groundwater-driven riparian whitesand valley. The DOM imprint of the riparian system was dominated by N-containing formulae potentially linked to poor drainage and anoxic conditions; this signature overprinted high-molecular weight aliphatic DOM signals of adjacent DOC-poor groundwater derived from elevated plateau areas. DOC from the dry Campina site was on average 6 years older (cal. ^{14}C age), but in general, exported DOC was young (< 11 years cal. ^{14}C age). The shared (putative tannic) markers of the dry Campina site and sets of Rio Negro samples showed strong overlap with previously reported molecular markers of low-pH environments, in particular forest soils. Their origin could be fungal but also resembled

recently reported sets of molecular formulae indicating preferential sorption to (and subsequent release from) iron oxides. These markers could thus indicate ongoing podsolization processes within the Rio Negro watershed. Our results demonstrate the decoupling of water and carbon fluxes of two contrasting ecosystems on the molecular level and generated new hypotheses on regional and whole-watershed scale processes contributing to the basin-scale Rio Negro imprint. However, further field and mechanistic studies are needed to constrain spatiotemporal dynamics of this ecosystem imprint, and to reveal marker identity.

In **chapter 5**, we develop a novel approach to identify potential ecosystem imprints of tannin-related polyphenolic plant metabolites in DOM stemming from shallow forest soil, where biological activity and SOM turnover is highest. We decipher complex chimeric mass spectra of soil DOM obtained by direct-injection electrospray ionization Orbitrap tandem mass spectrometry (DI-ESI-Orbitrap MS/MS) with collision-induced dissociation (CID). Pairwise exact mass differences (MDs) of all precursor and product ions yielded a tandem MS-based MD matrix that was subsequently matched against two lists of known MDs: 1) literature-known but non-indicative MDs, and 2) highly indicative MDs derived from tandem MS experiments using a set of 14 phenolic standard compounds. Additionally, we fragmented four isobaric precursor ion mixtures (IPIMs, m/z 241 301, 361 and 417) at three normalized collision energies (15, 20, 25) to analyze DOM fragmentation patterns and sensitivity. Our results reveal for the first time that MS fragmentation experiments help to identify DOM markers that relate to ecosystem processes. We show evidence for the presence of methyl radical losses indicative of lignin, polyol-equivalent losses indicative of glycosides, gallate-equivalent losses indicative of hydrolysable tannins, and series of MDs indicative of ring cleavage reactions of flavon-3-ols and flavan-3ols. We interpret the existence of these indicative higher-order MDs in DOM as a sign of a remaining source imprint of primary or recycled remains from plants, soil animals and microbes. However, the tannic identity of putative forest markers could not be proven, thus questioning the common routine of compound group annotations based on molecular formula alone. However, our approach resolved the contribution of MDs on the single precursor level, which added a layer of structural information to the frequently used Van Krevelen plot. We find that the matched MDs agree well with database structure suggestions. We also report on a significant correlation between the number of database hits and indicative MD matches of single precursors that yielded new insight into the underlying isomeric complexity of single DOM precursors. Altogether, these results open up new avenues to access structural information encrypted within complex mixtures and to link it to ecosystem processes.

Overall, the developed approaches and applications open up new ways to allow the analysis of overarching trends from ever growing numbers of FTMS DOM datasets. Dedicated instrument comparison studies and data collection tools are rare and need to be initiated in future. The comparison of published and new datasets and analysis of chimeric mass spectra has shown the high potential in generating new hypotheses of ecosystem marker origin and fate within soils, groundwater, streams, and river networks. These efforts have to be expanded to constrain sources and processes and eventually suggest marker structures by means of "semi"-targeted LC-tandem MS experiments (for example, by selected ion monitoring). Lastly, new approaches have to be developed to monitor the amount of ionized material and effects of ionization efficiency in order to improve quantification efforts by FTMS data. These efforts will be crucial to improve the understanding of the actual instrument response that is interpreted in a given biogeochemical context.

Zusammenfassung

Die Produktion und der Export von gelöstem organischen Material (DOM) steht in engem Zusammenhang mit Bodenprozessen, und die Gesellschaft ist in vielerlei Hinsicht auf die nachhaltige Funktion von Böden angewiesen (Küsel et al., 2016). Ein gutes Verständnis der Mechanismen, die den Umsatz der Bodenorganik und damit der DOM-Produktion steuern, ist daher maßgebend für ein ordnungsgemäßes Ökosystem-Management. Fortschritte in der instrumentellen Analytik – insbesondere der Kopplung sanfter Ionisationstechniken mit der hochauflösenden Fourier-Transformations-Massenspektrometrie (FTMS) – haben unsere Sicht auf diese allgegenwärtigen und hochkomplexen organischen Gemische revolutioniert. Dadurch ist die FTMS-Analytik trotz ihrer hohen Kosten zur gefragten Routinemethode avanciert, wenn es um die molekulare Aufklärung komplexer Gemische in der Biogeochemie geht (Minor et al., 2014; Smith et al., 2018; Zark and Dittmar, 2018). Im Vergleich zu den Fortschritten auf den Gebieten der Limnologie und der Ozeanographie ist unser Verständnis der Dynamik von DOM in Böden und im Grundwasser jedoch noch als gering einzuschätzen. Obwohl DOM heute als wertvoller Indikator und Integrator für rapide Ökosystemveränderungen angesehen wird (Roth et al., 2015, 2014), ist unser Wissensstand von molekularen Ökosystemabdrücken oder potenziellen neuen Prozessmarkern im DOM daher noch begrenzt. Im Hinblick auf diese Ausgangslage untersucht diese Arbeit die folgenden Punkte:

1) Bewertung der Eignung der Orbitrap als kostengünstigere Alternative zu modernen FT-ICR MS-Geräten für die Messung von terrestrischem DOM auf Basis eines Ansatzes zur Verbesserung der Reproduzierbarkeit zwischen DOM-Messungen der verschiedenen Geräte (**Kapitel 3**).

2) Anwendung der Orbitrap-Methode auf Proben aus dem Rio Negro-Becken: Extraktion von Ökosystem-Markern von DOM-reichen „Weißsand"-Gebieten, die im Zusammenhang mit der Produktion von Schwarzwässern stehen, mit anschließender Untersuchung der potenziellen Überschneidung mit Tannin-artigen molekularen Markern die für das Rio Negro-Becken beschrieben sind (**Kapitel 4**).

3) Entwicklung eines neuen Ansatzes um Ökosystemabdrücke in DOM mittels vergleichender Tandem-MS-Analysen an bekannten Tannin-artigen Polyphenolen zu untersuchen, mit anschließender Anwendung zur Identifizierung mutmaßlicher Tannin-artiger Marker im DOM einer exemplarischen Bodenwasserprobe aus einem Nadelwald (**Kapitel 5**).

In **Kapitel 3** demonstrieren wir einen Ansatz, der die Reproduzierbarkeit von FTMS-Spektren von komplexen DOM-Proben zwischen Messinstrumenten deutlich verbessert und zeigen weiterhin, dass die Orbitrap für die Analyse von DOM entlang terrestrisch-aquatischer Gradienten geeignet ist. Wir optimierten das Messsignal der Orbitrap hierfür mit einer einfachen "one sample/ whole spectrum"-Methode, die die Instrument-spezifischen Unterschiede zwischen den DOM-Massenspektren effektiv minimierte. Im Anschluss wurde die Orbitrap mit einem etablierten 15 Tesla FT-ICR MS anhand eines diversen Probensatzes verglichen bezüglich 1) Ionenhäufigkeitsmustern, 2) differentiellen Effekten des DOM-Typs auf den Informationsverlust und 3) abgeleiteten biogeochemischen Informationen. Wir konnten zeigen, dass die Orbitrap ähnliche Informationen wie das FT-ICR MS liefert, insbesondere für Ionen mit Massen < 400 *m/z*, und dass dies hauptsächlich durch das Auflösungsvermögen im höheren Massenbereich und nicht durch die Empfindlichkeit bedingt ist. Terrestrische Ökosysteme, die sich durch den Input von Pflanzenmaterial auszeichnen, z.B. Böden, Moore, Seen und Flüsse, zeigten bemerkenswert niedrige durchschnittliche molekulare Massen, sodass sie sich besser für Orbitrap-Messungen eigneten. Die zusätzlichen Informationen, die mittels FT-ICR MS gewonnen wurden, waren daher am höchsten in höher-massigen und Heteroatom-reicheren (Atome N, S, und P) Proben, die sich eher durch das Recycling von organischem Material auszeichnen, wie z.B. im Grundwasser und der Tiefsee. In diesen Proben detektierte die FT-ICR MS maximal +37% Summenformeln und +11% Signalintensität. Der gesamte Informationsgehalt, der mit multivariaten statistischen Methoden analysiert wurde, war jedoch für beide Datensätze vergleichbar. Die aus den Massenspektren abgeleiteten biogeochemischen Trends, wie z.B. die Abnahme der DOM-Aromatizität während des Transports durch verschiedene Ökosysteme, wurden von beiden Instrumenten gleichermaßen erfasst. Die Orbitrap eignet sich daher im besonderen Maße für anspruchsvolle Analysen an terrestrischem DOM mit Hinblick auf Komplexität, Herkunft und Umsatz. Weitere Arbeiten sind erforderlich, um den effektiv ionisierten Anteil an DOM abzuschätzen, um verbesserte quantitative Schätzungen auf der Grundlage von FTMS-Daten zu ermöglichen.

In **Kapitel 4** liefern wir molekulare Hinweise die die Hypothese bestätigen, dass Weißsand-Gebiete des Rio Negro-Beckens maßgeblichen Einfluss auf die Produktion von Schwarzwässern haben, und unterschiedliche molekulare Abdrücke innerhalb vom exportierten DOM hinterlassen, die eine gute Übereinstimmung mit hochoxidierten aromatischen (d.h. potenziell Tannin-haltigen) Rio-Negro-Markern zeigen. Wir verglichen hierzu standortspezifische molekulare Fingerabdrücke von Grundwasser-Transekten in zwei

ökohydrologisch unterschiedlichen Quellgebieten mit veröffentlichten DOM-Datensätzen des unteren Rio Negro, in den beide entwässern. Wasserisotope (^{18}O, ^2H) und Kohlenstoffisotope des DOC (^{13}C, ^{14}C) wurden gemessen um die Herkunft des Wassers und den Ursprung und das Alter des DOC abzuschätzen. DOM wurde mittels Orbitrap analysiert und auf Basis multivariater Statistik verglichen. Wir konnten zeigen, dass das trockenere, eher niederschlags-beeinflusste Weißsand-Gebiet (mit typischem *Campina*-Wald) stärkere Übereinstimmung mit Tannin-artigen und Rio Negro-spezifischen Strukturen zeigte, trotz der geschätzt zweifach geringeren Abflussmenge (aber 2-3 mal höherer DOC-Export) des Ökosystems im Vergleich zu einer nahegelegenen vernässten, Grundwasser-beeinflussten Weißsand-Talaue. Der DOM-Abdruck des Auensystems wurde von N-haltigen Formeln dominiert, die möglicherweise mit schlechter Entwässerung und anoxischen Bedingungen zusammenhängen. Die Talauen-Signatur überdeckte zudem die größtenteils aliphatische DOM-Signatur des DOC-armen Grundwassers aus angrenzenden, erhöhten Plateaugebieten. Der DOC vom trockenen Standort war im Durchschnitt 6 Jahre älter (kalibriertes ^{14}C-Alter), aber im Allgemeinen war der exportierte DOC jung (< 11 Jahre kalibriertes ^{14}C-Alter). Die gemeinsamen (vermeintlich Tannin-artigen) Marker des trockenen *Campina*-Systems und der Probensätze vom unteren Rio Negro zeigten starke Überschneidungen mit zuvor beschriebenen Markern aus Standorten mit saurem pH, insbesondere Waldböden. Ihr Ursprung könnte auf Pilze zurückführbar sein, ähnelt in den molekularen Eigenschaften aber wiederholt beschriebenen DOM-Signaturen, die mit der Sorption/ Desorption durch Eisenoxide in Verbindung gebracht werden. Diese Marker könnten somit auch auf laufende Podsolierungs-Prozesse innerhalb des Rio Negro-Beckens hinweisen. Unsere Ergebnisse demonstrieren die Entkopplung von Wasser- und Kohlenstoffflüssen anhand von zwei deutlich verschiedenen Weißsand-Ökosystemen auf der molekularen Ebene. Weitere Feldstudien und mechanistische Studien sind erforderlich, um die räumlich-zeitliche Dynamik des Ökosystem-Fingerabdrucks abzuschätzen und die Identität und Herkunft der Marker aufzuklären.

In **Kapitel 5** entwickeln wir einen neuartigen Ansatz zur Identifizierung potenzieller Ökosystemabdrücke von Tannin-artigen polyphenolischen Pflanzenmetaboliten im DOM aus flachen Waldböden, wo die biologische Aktivität und der SOM-Umsatz am höchsten sind. Wir entschlüsseln komplexe (sogenannte „chimäre") Massenspektren von isolierten DOM-Massen, die durch Direktinjektions-Elektrospray-Ionisations-Orbitrap Tandem-Massenspektrometrie (DI-ESI-Orbitrap MS/MS) mit kollisionsinduzierter Dissoziation (CID) erhalten wurden. Paarweise exakte Massenunterschiede (MDs) aller Vorläufer- und Produktionen (precursor/

product ion) ergaben hierbei eine Tandem-MS-basierte MD-Matrix, die anschließend mit zwei Listen bekannter MDs verglichen wurde: 1) Literatur-bekannte, aber unindikative MDs und 2) hoch-indikative MDs, die aus Tandem-MS-Experimenten mit einem Satz von 14 bekannten phenolischen Standardverbindungen abgeleitet wurden. Zusätzlich wurden vier isobare Vorläufer-Ionengemische (IPIMs, m/z 241 301, 361 und 417) bei drei normierten Kollisionsenergien (15, 20, 25) fragmentiert, um die Variabilität der DOM-Fragmentierungsmuster und Empfindlichkeit zu analysieren. Unsere Ergebnisse zeigen erstmalig dass MS-Fragmentierungsexperimente helfen können Ökosystem-relevante DOM-Marker zu identifizieren. Wir detektieren Methylradikalverluste, die auf Lignin-Verbindungen hinweisen, Polyol-äquivalente Verluste, die auf Glykoside hinweisen, Gallat-äquivalente Verluste, die auf hydrolysierbare Tannine hinweisen, und eine Reihe von MDs, die auf Ringspaltungsreaktionen von Flavon-3-olen und Flavan-3-olen hinweisen. Wir interpretieren die Existenz dieser indikativen MDs in DOM als ein Zeichen für einen verbleibenden Abdruck von primären oder recycelten Überresten von Pflanzen, Bodenfauna und Mikroben. Die Tannin-artige Identität von vermeintlichen Waldmarkern konnte jedoch nicht nachgewiesen werden, sodass die gängige Routine der Identifizierung von Verbindungsgruppen allein auf Grundlage der Summenformel infrage gestellt werden muss. Unser Ansatz erlaubt die Zuordnung von MDs auf der Ebene der einzelnen Vorläufer (precursor ions), die Strukturinformationen zum häufig verwendeten Van Krevelen-Plot hinzufügten. Wir stellten auch fest, dass die MD-Zuweisungen gut mit den Vorschlägen in einer Datenbankstruktur übereinstimmten. Daraus ergab sich auch eine signifikante Korrelation zwischen der Anzahl der Datenbanktreffer (Strukturvorschläge) und der Anzahl indikativer zugewiesener MDs von Vorläufer-Ionen. Dieser Zusammenhang erbrachte neue Erkenntnisse über die zugrundeliegende isomere Komplexität einzelner Vorläufer-Ionen im DOM. Insgesamt eröffnen diese Ergebnisse neue Wege um Strukturinformationen, die in komplexen Gemischen verschlüsselt sind, aufzudecken und sie mit Ökosystemprozessen in Verbindung zu bringen.

Insgesamt eröffnen die entwickelten Ansätze und präsentierten Anwendungen neue Wege, um die Analyse übergreifender Trends aus einer ständig wachsenden Anzahl von FTMS-DOM-Datensätzen zu ermöglichen. Ringversuche und dezidierte DOM-Datenbanken fehlen derzeit und müssen in Zukunft weiter gefördert werden. Der Vergleich von veröffentlichten und neuen Datensätzen und die Analyse von chimären Massenspektren hat großes Potenzial, um neue Hypothesen über die Herkunft und das Schicksal von Ökosystemmarkern in Böden, Grundwasser, Bächen und Flussnetzen zu generieren, die immer detailliertere Fragestellungen

ermöglichen. Diese Bemühungen sollten zur „Mustererkennung" von spezifischen Quellen und Prozessen führen und schließlich die Ableitung von Marker-Strukturen mittels gezielter LC-MS^n-Experimente (z.B. durch selected ion monitoring, SIM) ermöglichen. Schließlich müssen neue Ansätze entwickelt werden, um die Menge des ionisierten Materials und die Auswirkungen der Ionisationseffizienz zu erfassen, um die Quantifizierung mittels FTMS-Daten zu verbessern. Daraus könnte sich auf lange Sicht ein verbessertes Verständnis des DOM-Signals ergeben, was auch die Interpretationsmöglichkeiten im biogeochemischen Kontext erweitern würde.

Bibliography

Adamczyk, B., Adamczyk, S., Smolander, A., Kitunen, V., Simon, J., 2018. Plant secondary metabolites - missing pieces in the soil organic matter puzzle of boreal forests. Soil Syst. 2, 2. doi:10.3390/soils2010002

Addiscott, T.M., 2010. Soil mineralization: An emergent process? Geoderma 160, 31–35. doi:10.1016/j.geoderma.2010.03.016

Aiken, G.R., McKnight, D.M., Thorn, K.A., Thurman, E.M., 1992. Isolation of hydrophilic organic acids from water using nonionic macroporous resins. Org. Geochem. 18, 567–573. doi:10.1016/0146-6380(92)90119-I

Alvarez-Cobelas, M., Angeler, D.G., Sánchez-Carrillo, S., Almendros, G., 2012. A worldwide view of organic carbon export from catchments. Biogeochemistry 107, 275–293. doi:10.1007/s10533-010-9553-z

Aragão, L.E.O.C., Anderson, L.O., Fonseca, M.G., Rosan, T.M., Vedovato, L.B., Wagner, F.H., Silva, C.V.J., Junior, C.H.L.S., Arai, E., Aguiar, A.P., Barlow, J., Berenguer, E., Deeter, M.N., Domingues, L.G., Gatti, L., Gloor, M., Malhi, Y., Marengo, J.A., Miller, J.B., Phillips, O.L., Saatchi, S., 2018. 21st Century drought-related fires counteract the decline of Amazon deforestation carbon emissions. Nat. Commun. 9, 536. doi:10.1038/s41467-017-02771-y

Arakawa, N., Aluwihare, L., 2015. Direct identification of diverse alicyclic terpenoids in suwannee river fulvic acid. Environ. Sci. Technol. 49, 4097–4105. doi:10.1021/es5055176

Arakawa, N., Aluwihare, L.I., Simpson, A.J., Soong, R., Stephens, B.M., Lane-Coplen, D., 2017. Carotenoids are the likely precursor of a significant fraction of marine dissolved organic matter. Sci. Adv. 3, e1602976. doi:10.1126/sciadv.1602976

Arnosti, C., Reintjes, G., Amann, R., 2018. A mechanistic microbial underpinning for the size-reactivity continuum of dissolved organic carbon degradation. Mar. Chem. 206, 93–99. doi:10.1016/j.marchem.2018.09.008

Ashley, G.M., 1998. Where are we headed? "Soft" rock research into the new millenium. Geol. Soc. Am., Abstr. 30 148.

Aufdenkampe, A.K., Hedges, J.I., Richey, J.E., Krusche, A. V., Llerena, C.A., 2001. Sorptive fractionation of dissolved organic nitrogen and amino acids onto fine sediments within the Amazon Basin. Limnol. Oceanogr. 46, 1921–1935. doi:10.4319/lo.2001.46.8.1921

Aufdenkampe, A.K., Mayorga, E., Raymond, P.A., Melack, J.M., Doney, S.C., Alin, S.R., Aalto, R.E., Yoo, K., 2011. Riverine coupling of biogeochemical cycles between land, oceans, and atmosphere. Front. Ecol. Environ. 9, 53–60. doi:10.1890/100014

Austin, A.T., Méndez, M.S., Ballaré, C.L., 2016. Photodegradation alleviates the lignin bottleneck for carbon turnover in terrestrial ecosystems. Proc. Natl. Acad. Sci. 113, 4392–4397. doi:10.1073/pnas.1516157113

Aye, N.S., Butterly, C.R., Sale, P.W.G., Tang, C., 2018. Interactive effects of initial pH and nitrogen status on soil organic carbon priming by glucose and lignocellulose. Soil Biol. Biochem. 123, 33–44. doi:10.1016/j.soilbio.2018.04.027

Bailey, V.L., Smith, A.P., Tfaily, M., Fansler, S.J., Bond-Lamberty, B., 2017. Differences in soluble organic carbon chemistry in pore waters sampled from different pore size domains. Soil Biol. Biochem. 107, 133–143. doi:10.1016/j.soilbio.2016.11.025

Baldrian, P., Šnaijdr, J., 2011. Lignocellulose-degrading enzymes in soils, in: Shukla, G., Varma, A. (Eds.), Soil

Enzymology. Springer, Heidelberg, pp. 167–186.

Baldrian, P., Štursová, M., 2011. Enzymes in forest soils, in: Shukla, G., Varma, A. (Eds.), Soil Enzymology. Springer, Heidelberg, pp. 61–74.

Balesdent, J., Basile-Doelsch, I., Chadoeuf, J., Cornu, S., Derrien, D., Fekiacova, Z., Hatté, C., 2018. Atmosphere - soil carbon transfer as a function of soil depth. Nature 559, 599–602. doi:10.1038/s41586-018-0328-3

Baluha, D.R., Blough, N. V., Del Vecchio, R., 2013. Selective mass labeling for linking the optical properties of chromophoric dissolved organic matter to structure and composition via ultrahigh resolution electrospray ionization mass spectrometry. Environ. Sci. Technol. 47, 9891–9897. doi:10.1021/es402400j

Bar-On, Y.M., Phillips, R., Milo, R., 2018. The biomass distribution on Earth. Proc. Natl. Acad. Sci. 115, 6506–6511. doi:10.1073/pnas.1711842115

Bardy, M., Derenne, S., Allard, T., Benedetti, M.F., Fritsch, E., 2011. Podzolisation and exportation of organic matter in black waters of the Rio Negro (upper Amazon basin, Brazil). Biogeochemistry 106, 71–88. doi:10.1007/s10533-010-9564-9

Barré, P., Quénéa, K., Vidal, A., Cécillon, L., Christensen, B.T., Kätterer, T., Macdonald, A., Petit, L., Plante, A.F., van Oort, F., Chenu, C., 2018. Microbial and plant-derived compounds both contribute to persistent soil organic carbon in temperate soils. Biogeochemistry 140, 81–92. doi:10.1007/s10533-018-0475-5

Baumeister, T.U.H., Ueberschaar, N., Pohnert, G., 2018. Gas-phase chemistry in the GC Orbitrap mass spectrometer. J. Am. Soc. Mass Spectrom. doi:10.1007/s13361-018-2117-5

Baveye, P.C., Otten, W., Kravchenko, A., Balseiro-Romero, M., Beckers, É., Chalhoub, M., Darnault, C., Eickhorst, T., Garnier, P., Hapca, S., Kiranyaz, S., Monga, O., Mueller, C.W., Nunan, N., Pot, V., Schlüter, S., Schmidt, H., Vogel, H.J., 2018. Emergent properties of microbial activity in heterogeneous soil microenvironments: Different research approaches are slowly converging, yet major challenges remain. Front. Microbiol. 9, 1929. doi:10.3389/fmicb.2018.01929

Bayat, P., Lesage, D., Cole, R.B., 2018. Low-energy collision-induced dissociation (low-energy CID), collision-induced dissociation (CID), and higher energy collision dissociation (HCD) mass spectrometry for structural elucidation of saccharides and clarification of their dissolution mechanism i. J. Mass Spectrom. 53, 705–716. doi:10.1002/jms.4205

Bell, N.G.A., Michalchuk, A.A.L., Blackburn, J.W.T., Graham, M.C., Uhrín, D., 2015. Isotope-Filtered 4D NMR Spectroscopy for Structure Determination of Humic Substances. Angew. Chemie - Int. Ed. 54, 8382–8385. doi:10.1002/anie.201503321

Benk, S.A., Li, Y., Roth, V.-N., Gleixner, G., 2018. Lignin Dimers as Potential Markers for 14C-young Terrestrial Dissolved Organic Matter in the Critical Zone. Front. Earth Sci. 6, 1–9. doi:10.3389/feart.2018.00168

Bianchi, T.S., Canuel, E.A., 2011. Chemical biomarkers in aquatic ecosystems, 1st ed. Princeton University Press, Princeton.

Bird, J.A., Torn, M.S., 2006. Fine roots vs. needles: A comparison of 13C and 15N dynamics in a ponderosa pine forest soil. Biogeochemistry 79, 361–382. doi:10.1007/s10533-005-5632-y

Blume, H.-P., Brümmer, G.W., Horn, R., Kandeler, E., Kögel-Knabner, I., Kretzschmar, R., Stahr, K., Wilke, B.M., 2016. Scheffer/ Schachtschabel Lehrbuch der Bodenkunde.

Böcker, S., Dührkop, K., 2016. Fragmentation trees reloaded. J. Cheminform. 8, 5. doi:10.1007/978-3-319-16706-0_10

Boiteau, R.M., Fansler, S.J., Farris, Y., Shaw, J.B., Koppenaal, D.W., Pasa-Tolic, L., Jansson, J.K., 2019.

Siderophore profiling of co-habitating soil bacteria by ultra-high resolution mass spectrometry. Metallomics 11, 166–175. doi:10.1039/c8mt00252e

Borer, B., Tecon, R., Or, D., 2018. Spatial organization of bacterial populations in response to oxygen and carbon counter-gradients in pore networks. Nat. Commun. 9, 769. doi:10.1038/s41467-018-03187-y

Borken, W., Ahrens, B., Schulz, C., Zimmermann, L., 2011. Site-to-site variability and temporal trends of DOC concentrations and fluxes in temperate forest soils. Glob. Chang. Biol. 17, 2428–2443. doi:10.1111/j.1365-2486.2011.02390.x

Bowen, G.J., Revenaugh, J., 2003. Interpolating the isotopic composition of modern meteoric precipitation. Water Resour. Res. 39, 1–13. doi:10.1029/2003WR002086

Boye, K., Noël, V., Tfaily, M.M., Bone, S.E., Williams, K.H., Bargar, J.R., Fendorf, S., 2017. Thermodynamically controlled preservation of organic carbon in floodplains. Nat. Geosci. 10, 415–419. doi:10.1038/ngeo2940

Brantley, S.L., Goldhaber, M.B., Ragnarsdottir, K.V., 2007. Crossing disciplines and scales to understand the critical zone. Elements 3, 307–314.

Bravard, S., Righi, D., 1990. Podzols in Amazonia. Catena 17, 461–475.

Bronk Ramsey, C., 2009. Bayesian analysis of radiocarbon dates. Radiocarbon 51, 337–360.

Brown, T.A., Jackson, B.A., Bythell, B.J., Stenson, A.C., 2016. Benefits of multidimensional fractionation for the study and characterization of natural organic matter. J. Chromatogr. A 1470, 84–96. doi:10.1016/j.chroma.2016.10.005

Bünemann, E.K., Bongiorno, G., Bai, Z., Creamer, R.E., De Deyn, G., de Goede, R., Fleskens, L., Geissen, V., Kuyper, T.W., Mäder, P., Pulleman, M., Sukkel, W., van Groenigen, J.W., Brussaard, L., 2018. Soil quality – A critical review. Soil Biol. Biochem. 120, 105–125. doi:10.1016/j.soilbio.2018.01.030

Cameron, L., Earley, S., 2015. The ecosystem - Movements, connections, tensions and translations. Geoforum 65, 473–481. doi:10.1016/j.geoforum.2015.03.015

Camino-Serrano, M., Graf Pannatier, E., Vicca, S., Luyssaert, S., Jonard, M., Ciais, P., Guenet, B., Gielen, B., Peñuelas, J., Sardans, J., Waldner, P., Etzold, S., Cecchini, G., Clarke, N., GaliÄ, Z., Gandois, L., Hansen, K., Johnson, J., Klinck, U., Lachmanová, Z., Lindroos, A.J., Meesenburg, H., Nieminen, T.M., Sanders, T.G.M., Sawicka, K., Seidling, W., Thimonier, A., Vanguelova, E., Verstraeten, A., Vesterdal, L., Janssens, I.A., 2016. Trends in soil solution dissolved organic carbon (DOC) concentrations across European forests. Biogeosciences 13, 5567–5585. doi:10.5194/bg-13-5567-2016

Cao, D., Lv, J., Geng, F., Rao, Z., Niu, H., Shi, Y., Cai, Y., Kang, Y., 2016. Ion Accumulation Time Dependent Molecular Characterization of Natural Organic Matter Using Electrospray Ionization-Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. Anal. Chem. 88, 12210–12218. doi:10.1021/acs.analchem.6b03198

Capley, E.N., Tipton, J.D., Marshall, A.G., Stenson, A.C., 2010. Chromatographic reduction of isobaric and isomeric complexity of fulvic acids to enable multistage tandem mass spectral characterization. Anal. Chem. 82, 8194–8202. doi:10.1021/ac1016216

Carlson, C.A., Hansell, D.A., 2014. DOM Sources, Sinks, Reactivity, and Budgets, in: Carlson, C.A., Hansell, D.A. (Eds.), Biogeochemistry of Marine Dissolved Organic Matter. Elsevier B.V., Amsterdam, pp. 65–126. doi:10.1016/B978-0-12-405940-5.00003-0

Castellano, M.J., Mueller, K.E., Olk, D.C., Sawyer, J.E., Six, J., 2015. Integrating plant litter quality, soil organic matter stabilization, and the carbon saturation concept. Glob. Chang. Biol. 21, 3200–3209. doi:10.1111/gcb.12982

Cerri, C.C., Volkoff, B., Andreaux, F., 1991. Nature and behaviour of organic matter in soils under natural forest, and after deforestation, burning and cultivation, near Manaus. For. Ecol. Manage. 38, 247–257. doi:10.1016/0378-1127(91)90146-M

Cerri, C.E.P., Easter, M., Paustian, K., Killian, K., Coleman, K., Bernoux, M., Falloon, P., Powlson, D.S., Batjes, N.H., Milne, E., Cerri, C.C., 2007. Predicted soil organic carbon stocks and changes in the Brazilian Amazon between 2000 and 2030. Agric. Ecosyst. Environ. 122, 58–72. doi:10.1016/j.agee.2007.01.008

Chambers, M.C., Maclean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., Hoff, K., Kessner, D., Tasman, N., Shulman, N., Frewen, B., Baker, T. a, Brusniak, M.-Y., Paulse, C., Creasy, D., Flashner, L., Kani, K., Moulding, C., Seymour, S.L., Nuwaysir, L.M., Lefebvre, B., Kuhlmann, F., Roark, J., Rainer, P., Detlev, S., Hemenway, T., Huhmer, A., Langridge, J., Connolly, B., Chadick, T., Holly, K., Eckels, J., Deutsch, E.W., Moritz, R.L., Katz, J.E., Agus, D.B., MacCoss, M., Tabb, D.L., Mallick, P., 2012. A cross-platform toolkit for mass spectrometry and proteomics. Nat. Biotechnol. 30, 918–920. doi:10.1038/nbt.2377

Chassagne, F., Cabanac, G., Hubert, G., David, B., Marti, G., 2019. The landscape of natural product diversity and their pharmacological relevance from a focus on the Dictionary of Natural Products®. Phytochem. Rev. 1–22. doi:10.1007/s11101-019-09606-2

Chassé, A.W., Ohno, T., Higgins, S.R., Amirbahman, A., Yildirim, N., Parr, T.B., 2015. Chemical force spectroscopy evidence supporting the layer-by-layer model of organic matter binding to iron (oxy)hydroxide mineral surfaces. Environ. Sci. Technol. 49, 9733–9741. doi:10.1021/acs.est.5b01877

Chen, H., Abdulla, H.A.N., Sanders, R.L., Myneni, S.C.B., Mopper, K., Hatcher, P.G., 2014. Production of black carbon-like and aliphatic molecules from terrestrial dissolved organic matter in the presence of sunlight and iron. Environ. Sci. Technol. Lett. 1, 399–404. doi:10.1021/ez5002598

Chen, M., Kim, S., Park, J.E., Jung, H.J., Hur, J., 2016. Structural and compositional changes of dissolved organic matter upon solid-phase extraction tracked by multiple analytical tools. Anal. Bioanal. Chem. 408, 6249–6258. doi:10.1007/s00216-016-9728-0

Cortés-Francisco, N., Caixach, J., 2015. Fragmentation studies for the structural characterization of marine dissolved organic matter. Anal. Bioanal. Chem. 407, 2455–2462. doi:10.1007/s00216-015-8499-3

Cortés-Francisco, N., Flores, C., Moyano, E., Caixach, J., 2011. Accurate mass measurements and ultrahigh-resolution: Evaluation of different mass spectrometers for daily routine analysis of small molecules in negative electrospray ionization mode. Anal. Bioanal. Chem. 400, 3595–3606. doi:10.1007/s00216-011-5046-8

Cotrufo, M.F., Soong, J.L., Horton, A.J., Campbell, E.E., Haddix, M.L., Wall, D.H., Parton, W.J., 2015. Formation of soil organic matter via biochemical and physical pathways of litter mass loss. Nat. Geosci. 8, 776–779. doi:10.1038/ngeo2520

Coward, E.K., Ohno, T., Plante, A.F., 2018. Adsorption and molecular fractionation of dissolved organic matter on iron-bearing mineral matrices of varying crystallinity. Environ. Sci. Technol. 52, 1036–1044. doi:10.1021/acs.est.7b04953

Creed, I.F., McKnight, D.M., Pellerin, B.A., Green, M.B., Bergamaschi, B.A., Aiken, G.R., Burns, D.A., Findlay, S.E.G., Shanley, J.B., Striegl, R.G., Aulenbach, B.T., Clow, D.W., Laudon, H., McGlynn, B.L., McGuire, K.J., Smith, R.A., Stackpoole, S.M., 2015. The river as a chemostat: fresh perspectives on dissolved organic matter flowing down the river continuum. Can. J. Fish. Aquat. Sci. 72, 1272–1285. doi:10.1139/cjfas-2014-0400

Creed, I.F., Sanford, S.E., Beall, F.D., Molot, L.A., Dillon, P.J., 2003. Cryptic wetlands: Integrating hidden wetlands in regression models of the export of dissolved organic carbon from forested landscapes. Hydrol. Process. 17, 3629–3648. doi:10.1002/hyp.1357

D'Andrilli, J., Cooper, W.T., Foreman, C.M., Marshall, A.G., 2015. An ultrahigh-resolution mass spectrometry index to estimate natural organic matter lability. Rapid Commun. Mass Spectrom. 29, 2385–2401. doi:10.1002/rcm.7400

D'Andrilli, J., Dittmar, T., Koch, B.P., Purcell, J.M., Marshall, A.G., Cooper, W.T., 2010. Comprehensive chracterization of marine dissolved organic matter by Fourier transform ion cyclotron resonsance mass spectrometry with electrospray and atmospheric pressure photoionization. Rapid Commun. Mass Spectrom. 24, 643–650. doi:10.1002/rcm

Davies, N.W., Sandron, S., Nesterenko, P., Paull, B., Wilson, R., Haddad, P., Shellie, R., Rojas, A., 2015. Comment on "Structural characterization of dissolved organic matter: a review of current techniques for isolation and analysis" by E. C. Minor, M. M. Swenson, B. M. Mattson, and A. R. Oyler, Environ. Sci.: Processes Impacts, 2014, 16 , 2064. Environ. Sci. Process. Impacts 17, 495. doi:10.1039/C4EM00631C

De Oliveira Marques, J.D., Luizão, F.J., Teixeira, W.G., Ferreira, S.J.F., 2012. Variações do Carbono Orgânico Dissolvido e de Atributos Físicos do Solo Sob Diferentes Sistemas de Uso da Terra na Amazônia Central. Rev. Bras. Cienc. do Solo 36, 611–622. doi:10.1590/S0100-06832012000200030

De Oliveira Marques, J.D., Luizão, F.J., Teixeira, W.G., Vitel, C.M., Marques, E.M. de A., 2016. Soil organic carbon, carbon stock and their relationships to physical attributes under forest soils in central Amazonia. Rev. Árvore 40, 197–208. doi:10.1590/0100-67622016000200002

Dean, J.F., Garnett, M.H., Spyrakos, E., Billett, M.F., 2019. The potential hidden age of dissolved organic carbon exported by peatland streams. J. Geophys. Res. Biogeosciences 124, 328–341. doi:10.1029/2018jg004650

Delgado-Baquerizo, M., Maestre, F.T., Reich, P.B., Jeffries, T.C., Gaitan, J.J., Encinar, D., Berdugo, M., Campbell, C.D., Singh, B.K., 2016. Microbial diversity drives multifunctionality in terrestrial ecosystems. Nat. Commun. 7, 10541. doi:10.1038/ncomms10541

Demarchi, L.O., Scudeller, V.V., Moura, L.C., Dias-Terceiro, R.G., LOPES, A., Wittmann, F.K., Piedade, M.T.F., 2018. Floristic composition, structure and soil-vegetation relations in three white-sand soil patches in central Amazonia. Acta Amaz. 48, 46–56. doi:10.1590/1809-4392201603523

Demenois, J., Rey, F., Ibanez, T., Stokes, A., Carriconde, F., 2018. Linkages between root traits, soil fungi and aggregate stability in tropical plant communities along a successional vegetation gradient. Plant Soil 424, 319–334. doi:10.1007/s11104-017-3529-x

Denisov, E., Damoc, E., Lange, O., Makarov, A., 2012. Orbitrap mass spectrometry with resolving powers above 1,000,000. Int. J. Mass Spectrom. 325–327, 80–85. doi:10.1016/j.ijms.2012.06.009

Dit Fouque, D.J., Maroto, A., Memboeuf, A., 2016. Purification and quantification of an isomeric compound in a mixture by collisional excitation in multistage mass spectrometry experiments. Anal. Chem. 88, 10821–10825. doi:10.1021/acs.analchem.6b03490

Dittmar, T., Koch, B., Hertkorn, N., Kattner, G., 2008. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. Limnol. Oceanogr. Methods 6, 230–235. doi:10.4319/lom.2008.6.230

Dittmar, T., Koch, B.P., 2006. Thermogenic organic matter dissolved in the abyssal ocean. Mar. Chem. 102, 208–217. doi:10.1016/j.marchem.2006.04.003

Dittmar, T., Stubbins, A., 2014. Dissolved Organic Matter in aquatic systems, in: Holland, H., Turekian, K., Falkowski, P.G., Freeman, K.H. (Eds.), Treatise on Geochemistry: Second Edition. Volume 12: Organic Geochemistry. Elsevier Ltd, Oxford, pp. 125–156.

Do Nascimento, N.R., Bueno, G.T., Fritsch, E., Herbillon, A.J., Allard, T., Melfi, A.J., Astolfo, R., Boucher, H., Li, Y., 2004. Podzolization as a deferralitization process: A study of an Acrisol-Podzol sequence derived

from Palaeozoic sandstones in the northern upper Amazon Basin. Eur. J. Soil Sci. 55, 523–538. doi:10.1111/j.1365-2389.2004.00616.x

Do Nascimento, N.R., Fritsch, E., Bueno, G.T., Bardy, M., Grimaldi, C., Melfi, A.J., 2008. Podzolization as a deferralitization process: Dynamics and chemistry of ground and surface waters in an Acrisol - Podzol sequence of the upper Amazon Basin. Eur. J. Soil Sci. 59, 911–924. doi:10.1111/j.1365-2389.2008.01049.x

Don, A., Rödenbeck, C., Gleixner, G., 2013. Unexpected control of soil carbon turnover by soil carbon concentration. Environ. Chem. Lett. 11, 407–413. doi:10.1007/s10311-013-0433-3

Dosskey, M.G., Bertsch, P.M., 1994. Forest sources and pathways of organic matter transport to a blackwater stream : A hydrologic approach. Biogeochemistry 24, 1–19.

Doupoux, C., Merdy, P., Montes, C.R., Nunan, N., José Melfi, A., José Ribeiro Pereira, O., Lucas, Y., 2017. Modelling the genesis of equatorial podzols: Age and implications for carbon fluxes. Biogeosciences 14, 2429–2440. doi:10.5194/bg-14-2429-2017

Drake, T.W., Guillemette, F., Hemingway, J.D., Chanton, J.P., Podgorski, D.C., Zimov, N.S., Spencer, R.G.M., 2018. The ephemeral signature of permafrost Carbon in an Arctic fluvial network. J. Geophys. Res. Biogeosciences 123, 1475–1485. doi:10.1029/2017JG004311

Dubinenkov, I., Flerus, R., Schmitt-Kopplin, P., Kattner, G., Koch, B.P., 2015. Origin-specific molecular signatures of dissolved organic matter in the Lena Delta. Biogeochemistry 123, 1–14. doi:10.1007/s10533-014-0049-0

Dührkop, K., Shen, H., Meusel, M., Rousu, J., Böcker, S., 2015. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. Proc. Natl. Acad. Sci. 112, 12580–12585. doi:10.1073/pnas.1509788112

Duncan, K.D., Volmer, D.A., Gill, C.G., Krogh, E.T., 2016. Rapid screening of carboxylic acids from waste and surface waters by ESI-MS/MS using Barium ion chemistry and on-line membrane sampling. J. Am. Soc. Mass Spectrom. 27, 443–450. doi:10.1007/s13361-015-1311-y

Eisenhauer, N., Bonn, A., A. Guerra, C., 2019. Recognizing the quiet extinction of invertebrates. Nat. Commun. 10, 50. doi:10.1038/s41467-018-07916-1

Engström, M.T., Pälijärvi, M., Salminen, J.P., 2015. Rapid fingerprint analysis of plant extracts for ellagitannins, gallic acid, and quinic acid derivatives and quercetin-, kaempferol- and myricetin-based flavonol glycosides by UPLC-QqQ-MS/MS. J. Agric. Food Chem. 63, 4068–4079. doi:10.1021/acs.jafc.5b00595

Ertel, J.R., Hedges, J.I., Devol, A.H., Richey, J.E., De Nazare Goes Ribeiro, M., 1986. Dissolved humic substances of the Amazon River system. Limnol. Oceanogr. 31, 739–754.

Fabre, N., Rustan, I., De Hoffmann, E., Quetin-Leclercq, J., 2001. Determination of flavone, flavonol, and flavanone aglycones by negative ion liquid chromatography electrospray ion trap mass spectrometry. J. Am. Soc. Mass Spectrom. 12, 707–715. doi:10.1016/S1044-0305(01)00226-4

Fatayer, S., Coppola, A.I., Schulz, F., Walker, B.D., Broek, T.A., Meyer, G., Druffel, E.R.M., McCarthy, M., Gross, L., 2018. Direct Visualization of Individual Aromatic Compound Structures in Low Molecular Weight Marine Dissolved Organic Carbon. Geophys. Res. Lett. 45, 5590–5598. doi:10.1029/2018GL077457

Fievre, A., Solouki, T., Marshall, A.G., Cooper, W.T., 1997. High-resolution Fourier transform ion cyclotron resonance mass spectrometry of humic and fulvic acids by laser desorption/ ionization and electrospray ionization. Energy & Fuels 11, 554–560. doi:10.1021/ef970005q

Filella, M., 2014. Understanding what we are measuring: Standards and quantification of natural organic matter. Water Res. 50, 287–293. doi:10.1016/j.watres.2013.12.015

Filser, J., Faber, J.H., Tiunov, A. V., Brussaard, L., Frouz, J., De Deyn, G., Uvarov, A. V., Berg, M.P., Lavelle, P., Loreau, M., Wall, D.H., Querner, P., Eijsackers, H., Jiménez, J.J., 2016. Soil fauna: Key to new carbon models. Soil 2, 565–582. doi:10.5194/soil-2-565-2016

Finstad, A.G., Andersen, T., Larsen, S., Tominaga, K., Blumentrath, S., de Wit, H.A., Tømmervik, H., Hessen, D.O., 2016. From greening to browning: Catchment vegetation development and reduced S-deposition promote organic carbon load on decadal time scales in Nordic lakes. Sci. Rep. 6, 31944. doi:10.1038/srep31944

Fischer, U.A., Carle, R., Kammerer, D.R., 2011. Identification and quantification of phenolic compounds from pomegranate (Punica granatum L.) peel, mesocarp, aril and differently produced juices by HPLC-DAD-ESI/MSn. Food Chem. 127, 807–821. doi:10.1016/j.foodchem.2010.12.156

Flerus, R., Koch, B.P., Schmitt-Kopplin, P., Witt, M., Kattner, G., 2011. Molecular level investigation of reactions between dissolved organic matter and extraction solvents using FT-ICR MS. Mar. Chem. 124, 100–107. doi:10.1016/j.marchem.2010.12.006

Flerus, R., Lechtenfeld, O.J., Koch, B.P., McCallister, S.L., Schmitt-Kopplin, P., Benner, R., Kaiser, K., Kattner, G., 2012. A molecular perspective on the ageing of marine dissolved organic matter. Biogeosciences 9, 1935–1955. doi:10.5194/bg-9-1935-2012

Follett, C.L., Repeta, D.J., Rothman, D.H., Xu, L., Santinelli, C., 2014. Hidden cycle of dissolved organic carbon in the deep ocean. Proc. Natl. Acad. Sci. 111, 16706–16711. doi:10.1073/pnas.1407445111

Frappart, F., Papa, F., Famiglietti, J.S., Prigent, C., Rossow, W.B., Seyler, F., 2008. Interannual variations of river water storage from a multiple satellite approach: A case study for the Rio Negro River basin. J. Geophys. Res. Atmos. 113, D21104. doi:10.1029/2007JD009438

Fritsch, E., Allard, T., Benedetti, M.F., Bardy, M., Do Nascimento, N.R., Li, Y., Calas, G., 2009. Organic complexation and translocation of ferric iron in podzols of the Negro River watershed. Separation of secondary Fe species from Al species. Geochim. Cosmochim. Acta 73, 1813–1825. doi:10.1016/j.gca.2009.01.008

Fritsch, E., Balan, E., Do Nascimento, N.R., Allard, T., Bardy, M., Bueno, G., Derenne, S., Melfi, A.J., Calas, G., 2011. Deciphering the weathering processes using environmental mineralogy and geochemistry: Towards an integrated model of laterite and podzol genesis in the Upper Amazon Basin. Comptes Rendus Geosci. 343, 188–198. doi:10.1016/j.crte.2010.11.002

Fröberg, M., Berggren, D., Bergkvist, B., Bryant, C., Mulder, J., 2006. Concentration and fluxes of dissolved organic carbon (DOC) in three Norway spruce stands along a climatic gradient in Sweden. Biogeochemistry 77, 1–23. doi:10.1007/s10533-004-0564-5

Furch, K., Junk, W.J., 1997. Physicochemical conditions in floodplains, in: Junk, W.J. (Ed.), The Central Amazon Floodplain. Ecology of a Pulsing System. Ecological Studies 126. Springer, Heidelberg, pp. 69–108.

Galaverna, R.S., Sampaio, P.T.B., Barata, L.E.S., Eberlin, M.N., Fidelis, C.H. V., 2015. Differentiation of two morphologically similar Amazonian Aniba species by mass spectrometry leaf fingerprinting. Anal. Methods 7, 1984–1990. doi:10.1039/c4ay02598a

Galindo, C., Del Nero, M., 2015. Chemical fractionation of a terrestrial humic acid upon sorption on alumina by high resolution mass spectrometry. RSC Adv. 5, 73058–73067. doi:10.1039/C5RA12091H

Galindo, C., Del Nero, M., 2014. Molecular level description of the sorptive fractionation of a fulvic acid on aluminum oxide using electrospray ionization fourier transform mass spectrometry. Environ. Sci. Technol. 48, 7401–7408. doi:10.1021/es501465h

Gallet, C., Pellissier, F., 1997. Phenolic compounds in natural solutions of a coniferous forest. J. Chemcial Ecol.

23, 2401–2412.

Gaspar, A., Kunenkov, E. V., Lock, R., Desor, M., Perminova, I., Schmitt-Kopplin, P., 2009. Combined utilization of ion mobility and ultra-high-resolution mass spectrometry to identify multiply charged constituents in natural organic matter. Rapid Commun. Mass Spectrom. 23, 683–688. doi:10.1002/rcm

Gehre, M., Geilmann, H., Richter, J., Werner, R.A., Brand, W.A., 2004. Continuous flow 2H/1H and 18O/ 16O analysis of water samples with dual inlet precision. Rapid Commun. Mass Spectrom. 18, 2650–2660. doi:10.1002/rcm.1672

Gleixner, G., 2013. Soil organic matter dynamics: A biological perspective derived from the use of compound-specific isotopes studies. Ecol. Res. 28, 683–695. doi:10.1007/s11284-012-1022-9

Gloor, E., 2016. Climate and the Amazonian Carbon Balance, in: Nagy, L., Forsberg, B.R., Artaxo, P. (Eds.), Interactions Between Biosphere, Atmosphere and Human Land Use in the Amazon Basin. Ecological Studies 227. Springer, Heidelberg, pp. 101–117.

Gologan, B., Takáts, Z., Alvarez, J., Wiseman, J.M., Talaty, N., Ouyang, Z., Cooks, R.G., 2004. Ion soft-landing into liquids: Protein identification, separation, and purification with retention of biological activity. J. Am. Soc. Mass Spectrom. 15, 1874–1884. doi:10.1016/j.jasms.2004.09.005

Gonsior, M., Valle, J., Schmitt-Kopplin, P., Hertkorn, N., Bastviken, D., Luek, J., Harir, M., Bastos, W., Enrich-Prast, A., 2016. Chemodiversity of dissolved organic matter in the Amazon Basin. Biogeosciences 13, 4279–4290. doi:10.5194/bg-13-4279-2016

Goulding, M., Carvalho, M.L., Ferreira, E.G., 1988. Blackwaters, in: Goulding, M., Carvalho, M.L., Ferreira, E.G. (Eds.), Rio Negro, Rich Life in Poor Water. Amazonian Diversity and Foodchain Ecology as Seen through Fish Communities. SPB Academic Publishing, The Hague, pp. 29–36. doi:10.1016/0169-5347(89)90067-0

Graven, H.D., 2015. Impact of fossil fuel emissions on atmospheric radiocarbon and various applications of radiocarbon over this century. Proc. Natl. Acad. Sci. 112, 9542–9545. doi:10.1073/pnas.1504467112

Green, N.W., Mcinnis, D., Hertkorn, N., Maurice, P.A., Perdue, M.E., 2014a. Suwannee River Natural Organic Matter : Isolation of the 2R101N Reference Sample by Reverse Osmosis. Environ. Eng. Sci. 32, 38–44. doi:10.1089/ees.2014.0284

Green, N.W., Perdue, E.M., Aiken, G.R., Butler, K.D., Chen, H., Dittmar, T., Niggemann, J., Stubbins, A., 2014b. An intercomparison of three methods for the large-scale isolation of oceanic dissolved organic matter. Mar. Chem. 161, 14–19. doi:10.1016/j.marchem.2014.01.012

Gross, C.D., Harrison, R.B., 2019. The case for digging deeper: Soil organic carbon storage, dynamics, and controls in our changing world. Soil Syst. 3, 28. doi:10.3390/soilsystems3020028

Gross, G.G., 2009. Biosynthesis of ellagitannins: old ideas and new solutions, in: Quideau, S. (Ed.), Chemistry and Biology of Ellagitannins. World Scientific, London, pp. 94–118. doi:10.1142/9789812797414_0003

Gross, J.H., 2011. Mass Spectrometry. A Textbook, 2nd ed. Springer, Heidelberg.

Gu, L., Kelm, M.A., Hammerstone, J.F., Beecher, G., Holden, J., Haytowitz, D., Prior, R.L., 2003. Screening of Foods Containing Proanthocyanidins and Their Structural Characterization Using LC-MS/MS and Thiolytic Degradation. J. Agric. Food Chem. 51, 7513–7521. doi:10.1021/jf034815d

Guggenberger, G., Kaiser, K., 2003. Dissolved organic matter in soil: Challenging the paradigm of sorptive preservation. Geoderma 113, 293–310. doi:10.1016/S0016-7061(02)00366-X

Guigue, J., Mathieu, O., Lévêque, J., Mounier, S., Laffont, R., Maron, P.A., Navarro, N., Chateau, C., Amiotte-

Suchet, P., Lucas, Y., 2014. A comparison of extraction procedures for water-extractable organic matter in soils. Eur. J. Soil Sci. 65, 520–530. doi:10.1111/ejss.12156

Guinoiseau, D., Bouchez, J., Gélabert, A., Louvat, P., Filizola, N., Benedetti, M.F., 2016. The geochemical filter of large river confluences. Chem. Geol. 441, 191–203. doi:10.1016/j.chemgeo.2016.08.009

Ha, J., Cho, E., Kim, S., 2017. Interpreting chemical structures of compounds in crude oil based on the tandem mass spectra of standard compounds obtained at the same normalized collision energy. Energy and Fuels 31, 6960–6967. doi:10.1021/acs.energyfuels.7b00882

Hagedorn, F., Schleppi, P., Waldner, P., Flühler, H., 2000. Export of dissolved organic carbon and nitrogen from Gleysol dominated catchments – the significance of water flow paths. Biogeochemistry 50, 137–161.

Häggi, C., Sawakuchi, A.O., Chiessi, C.M., Mulitza, S., Mollenhauer, G., Sawakuchi, H.O., Baker, P.A., Zabel, M., Schefuß, E., 2016. Origin, transport and deposition of leaf-wax biomarkers in the Amazon Basin and the adjacent Atlantic. Geochim. Cosmochim. Acta 192, 149–165. doi:10.1016/j.gca.2016.07.002

Hansman, R.L., Dittmar, T., Herndl, G.J., 2015. Conservation of dissolved organic matter molecular composition during mixing of the deep water masses of the northeast Atlantic Ocean. Mar. Chem. 177, 288–297. doi:10.1016/j.marchem.2015.06.001

Hartmann, A.C., Petras, D., Quinn, R.A., Protsyuk, I., Archer, F.I., Ransome, E., Williams, G.J., Bailey, B.A., Vermeij, M.J.A., Alexandrov, T., Dorrestein, P.C., Rohwer, F.L., 2017. Meta-mass shift chemical profiling of metabolomes from coral reefs. Proc. Natl. Acad. Sci. 114, 11685–11690. doi:10.1073/pnas.1710248114

Hawkes, J., Sjöberg, P.J.R., Bergquist, J., Tranvik, L., 2019. Complexity of dissolved organic matter in the molecular size dimension: insights from coupled size exclusion chromatography electrospray ionisation mass spectrometry. Faraday Discuss. Available Online. doi:10.1039/c8fd00222c

Hawkes, J.A., Dittmar, T., Patriarca, C., Tranvik, L., Bergquist, J., 2016. Evaluation of the Orbitrap Mass Spectrometer for the Molecular Fingerprinting Analysis of Natural Dissolved Organic Matter. Anal. Chem. 88, 7698–7704. doi:10.1021/acs.analchem.6b01624

Hawkes, J.A., Patriarca, C., Sjöberg, P.J.R., Tranvik, L.J., Bergquist, J., 2018a. Extreme isomeric complexity of dissolved organic matter found across aquatic environments. Limnol. Oceanogr. Lett. 3, 21–30. doi:10.1002/lol2.10064

Hawkes, J.A., Radoman, N., Bergquist, J., Wallin, M.B., Tranvik, L.J., Löfgren, S., 2018b. Regional diversity of complex dissolved organic matter across forested hemiboreal headwater streams. Sci. Rep. 8, 16060. doi:10.1038/s41598-018-34272-3

Heimann, M., Reichstein, M., 2008. Terrestrial ecosystem carbon dynamics and climate feedbacks. Nature 451, 289–292. doi:10.1038/nature06591

Heitkötter, J., Marschner, B., 2018. Is there anybody out there? substrate availability controls microbial activity outside of hotspots in subsoils. Soil Syst. 2, 35. doi:10.3390/soilsystems2020035

Hemmler, D., Roullier-Gall, C., Marshall, J.W., Rychlik, M., Taylor, A.J., Schmitt-Kopplin, P., 2017. Evolution of Complex Maillard Chemical Reactions, Resolved in Time. Sci. Rep. 7, 3227. doi:10.1038/s41598-017-03691-z

Hernes, P.J., Spencer, R.G.M., Dyda, R.Y., O'Geen, A.T., Dahlgren, R.A., 2017. The Genesis and Exodus of Vascular Plant DOM from an Oak Woodland Landscape. Front. Earth Sci. 5, 9. doi:10.3389/feart.2017.00009

Hertkorn, N., 2014. Environmental NMR: Solution-state methods, in: Simpson, M.J., Simpson, A.J. (Eds.), NMR Spectroscopy. A Versatile Tool for Environmental Research. John Wiley & Sons, Ltd., Chichester, pp. 3–30.

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

Hertkorn, N., Frommberger, M., Witt, M., Koch, B.P., Schmitt-Kopplin, P., Perdue, E.M., 2008. Natural Organic Matter and the Event Horizon of Mass Spectrometry. Anal. Chem. 80, 8908–8919.

Hertkorn, N., Harir, M., Cawley, K.M., Schmitt-Kopplin, P., Jaffé, R., 2016. Molecular characterization of dissolved organic matter from subtropical wetlands: A comparative study through the analysis of optical properties, NMR and FTICR/MS. Biogeosciences 13, 2257–2277. doi:10.5194/bgd-12-13711-2015

Hertkorn, N., Harir, M., Koch, B.P., Michalke, B., Schmitt-Kopplin, P., 2013. High-field NMR spectroscopy and FTICR mass spectrometry: Powerful discovery tools for the molecular level characterization of marine dissolved organic matter. Biogeosciences 10, 1583–1624. doi:10.5194/bg-10-1583-2013

Hertkorn, N., Ruecker, C., Meringer, M., Gugisch, R., Frommberger, M., Perdue, E.M., Witt, M., Schmitt-Kopplin, P., 2007. High-precision frequency measurements: Indispensable tools at the core of the molecular-level analysis of complex systems. Anal. Bioanal. Chem. 389, 1311–1327. doi:10.1007/s00216-007-1577-4

Herzsprung, P., Hertkorn, N., von Tümpling, W., Harir, M., Friese, K., Schmitt-Kopplin, P., 2014. Understanding molecular formula assignment of Fourier transform ion cyclotron resonance mass spectrometry data of natural organic matter from a chemical point of view. Anal. Bioanal. Chem. 406, 7977–7987. doi:10.1007/s00216-014-8249-y

Hicks Pries, C.E., Castanha, C., Porras, R.C., Torn, M.S., 2017. The whole-soil carbon flux in response to warming. Science (80-.). 355, 1420–1423. doi:10.1126/science.aal1319

Hockaday, W.C., Purcell, J.M., Marshall, A.G., Baldock, J.A., Hatcher, P.G., 2009. Electrospray and photoionization mass spectrometry for the characterization of organic matter in natural waters: a qualitative assessment. Limnol. Oceanogr. Methods 7, 81–95. doi:10.4319/lom.2009.7.81

Horbe, A.M.C., Lages, A. da S., Moquet, J.S., Santos, R.V., Seyler, P., Cochonneau, G., 2016. Geochemistry of organic-rich river waters in Amazonia: Insights on weathering processes of intertropical cratonic terrain. Appl. Geochemistry 65, 22–35. doi:10.1016/j.apgeochem.2015.10.007

Hua, Q., Barbetti, M., Rakowski, A.Z., 2013. Atmospheric radiocarbon for the period 1950–2010. Radiocarbon 55, 2059–2072.

Hughey, C.A., Hendrickson, C.L., Rodgers, R.P., Marshall, A.G., Qian, K., 2001. Kendrick mass defect spectrum: A compact visual analysis for ultrahigh-resolution broadband mass spectra. Anal. Chem. 73, 4676–4681. doi:10.1021/ac010560w

Hursh, A., Ballantyne, A., Cooper, L., Maneta, M., Kimball, J., Watts, J., 2017. The sensitivity of soil respiration to soil temperature, moisture, and carbon supply at the global scale. Glob. Chang. Biol. 23, 2090–2103. doi:10.1111/gcb.13489

Hutchins, R.H.S., Aukes, P., Schiff, S.L., Dittmar, T., Prairie, Y.T., del Giorgio, P.A., 2017. The Optical, Chemical, and Molecular Dissolved Organic Matter Succession Along a Boreal Soil-Stream-River Continuum. J. Geophys. Res. Biogeosciences 122, 2892–2908. doi:10.1002/2017JG004094

Ishida, D.A., Montes, C.R., Lucas, Y., Pereira, O.J.R., Merdy, P., Melfi, A.J., 2014. Genetic relationships between ferralsols, podzols and white kaolin in Amazonia. Eur. J. Soil Sci. 65, 706–717. doi:10.1111/ejss.12167

Jaffé, R., Yamashita, Y., Maie, N., Cooper, W.T., Dittmar, T., Dodds, W.K., Jones, J.B., Myoshi, T., Ortiz-Zayas, J.R., Podgorski, D.C., Watanabe, a., 2012. Dissolved Organic Matter in Headwater Streams: Compositional Variability across Climatic Regions of North America. Geochim. Cosmochim. Acta 94, 95–108.

doi:10.1016/j.gca.2012.06.031

James, J.N., Gross, C.D., Dwivedi, P., Myers, T., Santos, F., Bernardi, R., Fidalgo de Faria, M., Amaral Guerrini, I., Harrison, R., Butman, D., 2019. Land use change alters the radiocarbon age and composition of soil and water-soluble organic matter in the Brazilian Cerrado. Geoderma 345, 38–50. doi:10.1016/j.geoderma.2019.03.019

Johnson, G.E., Gunaratne, D., Laskin, J., 2016. Soft- and reactive landing of ions onto surfaces: Concepts and applications. Mass Spectrosc. Rev. 35, 439–479. doi:10.1002/mas

Juyal, A., Eickhorst, T., Falconer, R., Baveye, P.C., Spiers, A., Otten, W., 2018. Control of pore geometry in soil microcosms and its effect on the growth and spread of Pseudomonas and Bacillus sp. Front. Environ. Sci. 6, 73. doi:10.3389/fenvs.2018.00073

Kaiser, K., Guggenberger, G., Haumaier, L., Zech, W., 2001. Seasonal variations in the chemical composition of dissolved organic matter in organic forest floor layer leachates of old-growth Scots pine (Pinus sylvestris L.) and European beech (Fagus sylvatica L.) stands in northeastern Bavaria, Germany. Biogeochemistry 103–143. doi:10.1023/A:1010694032121

Kaiser, K., Kalbitz, K., 2012. Cycling downwards - dissolved organic matter in soils. Soil Biol. Biochem. 52, 29–32. doi:10.1016/j.soilbio.2012.04.002

Kaiser, M., Kleber, M., Berhe, A.A., 2015. How air-drying and rewetting modify soil organic matter characteristics: An assessment to improve data interpretation and inference. Soil Biol. Biochem. 80, 324–340. doi:10.1016/j.soilbio.2014.10.018

Kallenbach, C.M., Grandy, A., Frey, S.D., 2016. Direct evidence for microbial-derived soil organic matter formation and its ecophysiological controls. Nat. Commun. 7, 13630. doi:10.1038/ncomms13630

Kang, H., Kwon, M.J., Kim, S., Lee, S., Jones, T.G., Johncock, A.C., Haraguchi, A., Freeman, C., 2018. Biologically driven DOC release from peatlands during recovery from acidification. Nat. Commun. 9, 3807.

Kästner, M., Miltner, A., 2018. SOM and microbes - what is left from microbial life, The future of soil carbon: Its conservation and formation. Elsevier Inc. doi:10.1016/B978-0-12-811687-6.00005-5

Keeler, J., 2011. Understanding NMR spectroscopy, 2nd ed. John Wiley & Sons, Ltd., Chichester.

Kellerman, A.M., Dittmar, T., Kothawala, D.N., Tranvik, L.J., 2014. Chemodiversity of dissolved organic matter in lakes driven by climate and hydrology. Nat. Commun. 5, 3804. doi:10.1038/ncomms4804

Kellerman, A.M., Guillemette, F., Podgorski, D.C., Aiken, G.R., Butler, K.D., Spencer, R.G.M., 2018. Unifying Concepts Linking Dissolved Organic Matter Composition to Persistence in Aquatic Ecosystems. Environ. Sci. Technol. 52, 2538–2548. doi:10.1021/acs.est.7b05513

Kellerman, A.M., Kothawala, D.N., Dittmar, T., Tranvik, L.J., 2015. Persistence of dissolved organic matter in lakes related to its molecular characteristics. Nat. Geosci. 8. doi:10.1038/ngeo2440

Kew, W., Blackburn, J.W.T., Clarke, D.J., Uhrín, D., 2017. Interactive van Krevelen diagrams - Advanced visualisation of mass spectrometry data of complex mixtures. Rapid Commun. Mass Spectrom. 31, 658–662. doi:10.1002/rcm.7823

Kim, S., Kramer, R.W., Hatcher, P.G., 2003. Graphical Method for Analysis of Ultrahigh-Resolution Braodband mass spectra of Natural Organic Matter, the Van Krevelen diagram. Anal. Chem. 75, 5336–5344.

Kind, T., Fiehn, O., 2007. Seven golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. BMC Bioinformatics 8, 105–125. doi:10.1186/1471-2105-8-105

Kindler, R., Siemens, J., Kaiser, K., Walmsley, D.C., Bernhofer, C., Buchmann, N., Cellier, P., Lehuger, S., Jones, S.K., Skiba, U., Eugster, W., Ibrom, A., Kutsch, W., Osborne, B., Soussana, J.-F., Tefs, C., Moors, E., Heim, A., Saunders, M., Jones, M., Grünwald, T., Gleixner, G., Loubet, B., McKenzie, R., Pilegaard, K., Schmidt, M.W.I., Zeeman, M.J., Seyfferth, J., Larsen, K.S., Vowinckel, B., Klumpp, K., Schrumpf, M., Rebmann, C., Sutton, M.A., Kaupenjohann, M., 2010. Dissolved carbon leaching from soil is a crucial component of the net ecosystem carbon balance. Glob. Chang. Biol. 17, 1167–1185. doi:10.1111/j.1365-2486.2010.02282.x

Kleber, M., Lehmann, J., 2019. Humic substances extracted by alkali are invalid proxies for the dynamics and functions of organic matter in terrestrial and aquatic ecosystems. J. Environ. Qual. 48, 207–216. doi:10.2134/jeq2019.01.0036

Kleber, M., Sollins, P., Sutton, R., 2007. A conceptual model of organo-mineral interactions in soils : self-assembly of organic molecular fragments into zonal structures on mineral surfaces. Biogeochemistry 85, 9–24. doi:10.1007/s10533-007-9103-5

Kleidon, A., 2016. Thermodynamic Foundations of the Earth System, 1st ed. Cambridge University Press, Cambridge.

Klotzbücher, T., Kalbitz, K., Cerli, C., Hernes, P.J., Kaiser, K., 2016. Gone or just out of sight? the apparent disappearance of aromatic litter components in soils. Soil 2, 325–335. doi:10.5194/soil-2-325-2016

Koch, B.P., Dittmar, T., 2016. From mass to structure: An aromaticity index for high-resolution mass data of natural organic matter. Rapid Commun. Mass Spectrom. 30, 250. doi:10.1002/rcm.7433

Koch, B.P., Dittmar, T., 2006. From mass to structure: An aromaticity index for high-resolution mass data of natural organic matter. Rapid Commun. Mass Spectrom. 30, 250. doi:10.1002/rcm.2386

Koch, B.P., Dittmar, T., Witt, M., Kattner, G., 2007. Fundamentals of molecular formula assignment to ultrahigh resolution mass data of Natural Organic Matter. Anal. Chem. 79, 1758–1763. doi:10.1021/ac061949s

Kögel-Knabner, I., 2017. The macromolecular organic composition of plant and microbial residues as inputs to soil organic matter: Fourteen years on. Soil Biol. Biochem. 105, A3–A8. doi:10.1016/j.soilbio.2016.08.011

Kögel-Knabner, I., 2002. The macromolecular organic composition of plant and microbial residues as inputs to soil organic matter. Soil Biol. Biochem. 34, 139–162.

Kostyukevich, Y., Kononikhin, A., Zherebker, A., Popov, I., Perminova, I., Nikolaev, E., 2014. Enumeration of non-labile oxygen atoms in dissolved organic matter by use of 16O/18O exchange and Fourier transform ion-cyclotron resonance mass spectrometry. Anal. Bioanal. Chem. 406, 6655–6664. doi:10.1007/s00216-014-8097-9

Kothawala, D.N., Ji, X., Laudon, H., Ågren, A.M., Futter, M.N., Köhler, S.J., Tranvik, L.J., 2015. The relative influence of land cover, hydrology, and in-stream processing on the composition of dissolved organic matter in boreal streams. J. Geophys. Res. Biogeosciences 120, 1491–1505. doi:10.1002/2015JG002946

Kroll, J.H., Donahue, N.M., Jimenez, J.L., Kessler, S.H., Canagaratna, M.R., Wilson, K.R., Altieri, K.E., Mazzoleni, L.R., Wozniak, A.S., Bluhm, H., Mysak, E.R., Smith, J.D., Kolb, C.E., Worsnop, D.R., 2011. Carbon oxidation state as a metric for describing the chemistry of atmospheric organic aerosol. Nat. Chem. 3, 133–139. doi:10.1038/nchem.948

Krueger, J., Heitkötter, J., Leue, M., Schlüter, S., Vogel, H.J., Marschner, B., Bachmann, J., 2018. Coupling of interfacial soil properties and bio-hydrological processes: The flow cell concept. Ecohydrology 11, e2024. doi:10.1002/eco.2024

Kügler, S., Cooper, R.E., Wegner, C.E., Mohr, J.F., Wichard, T., Küsel, K., 2019. Iron-organic matter complexes accelerate microbial iron cycling in an iron-rich fen. Sci. Total Environ. 646, 972–988. doi:10.1016/j.scitotenv.2018.07.258

Kuhlisch, C., Pohnert, G., 2015. Metabolomics in chemical ecology. Nat. Prod. Rep. 32, 937–955. doi:10.1039/c5np00003c

Kuhnert, N., Dairpoosh, F., Yassin, G., Golon, A., Jaiswal, R., 2013. What is under the hump? Mass spectrometry based analysis of complex mixtures in processed food – lessons from the characterisation of black tea thearubigins, coffee melanoidines and caramel. Food Funct. 4, 1130. doi:10.1039/c3fo30385c

Kujawinski, E.B., 2011. The impact of microbial metabolism on marine dissolved organic matter. Ann. Rev. Mar. Sci. 3, 567–599. doi:10.1146/annurev-marine-120308-081003

Kujawinski, E.B., Behn, M.D., 2006. Automated analysis of electrospray ionization fourier transform ion cyclotron resonance mass spectra of natural organic matter. Anal. Chem. 78, 4363–4373. doi:10.1021/ac0600306

Kujawinski, E.B., Del Vecchio, R., Blough, N. V., Klein, G.C., Marshall, A.G., 2004. Probing molecular-level transformations of dissolved organic matter: insights on photochemical degradation and protozoan modification of DOM from electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. Mar. Chem. 92, 23–37. doi:10.1016/j.marchem.2004.06.038

Kujawinski, E.B., Longnecker, K., Blough, N. V., Vecchio, R. Del, Finlay, L., Kitner, J.B., Giovannoni, S.J., 2009. Identification of possible source markers in marine dissolved organic matter using ultrahigh resolution mass spectrometry. Geochim. Cosmochim. Acta 73, 4384–4399. doi:10.1016/j.gca.2009.04.033

Kunenkov, E. V., Kononikhin, A.S., Perminova, I. V., Hertkorn, N., Gaspar, A., Schmitt-kopplin, P., Popov, I.A., Garmash, A. V., Nikolaev, E.N., 2009. Total Mass Difference Statistics Algorithm : A New Approach to Identification of High-Mass Building Blocks in Electrospray Ionization Fourier Transform Ion Cyclotron Mass Spectrometry Data of Natural Organic Matter. Anal. Chem. 81, 10106–10115. doi:10.1021/ac901476u

Kunert, N., Aparecido, L.M.T., Wolff, S., Higuchi, N., Dos Santos, J., De Araujo, A.C., Trumbore, S., 2017. A revised hydrological model for the Central Amazon: The importance of emergent canopy trees in the forest water budget. Agric. For. Meteorol. 239, 47–57. doi:10.1016/j.agrformet.2017.03.002

Küsel, K., Totsche, K.U., Trumbore, S.E., Lehmann, R., Steinhäuser, C., Herrmann, M., 2016. How Deep Can Surface Signals Be Traced in the Critical Zone? Merging Biodiversity with Biogeochemistry Research in a Central German Muschelkalk Landscape. Front. Earth Sci. 4, 32. doi:10.3389/feart.2016.00032

Kuzyakov, Y., Blagodatskaya, E., 2015. Microbial hotspots and hot moments in soil: Concept & review. Soil Biol. Biochem. 83, 184–199. doi:10.1016/j.soilbio.2015.01.025

Lam, B., Baer, A., Alaee, M., Lefebvre, B., Moser, A., Williams, A., Simpson, A.J., 2007. Major structural components in freshwater dissolved organic matter. Environ. Sci. Technol. 41, 8240–8247. doi:10.1021/es0713072

Lambert, T., Bouillon, S., Darchambeau, F., Massicotte, P., Borges, A. V., 2016. Shift in the chemical composition of dissolved organic matter in the Congo River network. Biogeosciences 13, 5405–5420. doi:10.5194/bg-13-5405-2016

Lange, M., Eisenhauer, N., Sierra, C.A., Bessler, H., Engels, C., Griffiths, R.I., Mellado-Vázquez, P.G., Malik, A.A., Roy, J., Scheu, S., Steinbeiss, S., Thomson, B.C., Trumbore, S.E., Gleixner, G., 2015. Plant diversity increases soil microbial activity and soil carbon storage. Nat. Commun. 6, 6707. doi:10.1038/ncomms7707

Langmuir, C.H., Broecker, W., 2012. How to Build a Habitable Planet: The Story of Earth from the Big Bang to Humankind, Revised an. ed. Princeton University Press, Princeton.

Laudon, H., Sponseller, R.A., 2018. How landscape organization and scale shape catchment hydrology and biogeochemistry: insights from a long-term catchment study. Wiley Interdiscip. Rev. Water 5, e1265. doi:10.1002/wat2.1265

Lavelle, P., Spain, A., Blouin, M., Brown, G., Decaëns, T., Grimaldi, M., Jiménez, J.J., McKey, D., Mathieu, J., Valesquez, E., Zangerlé, A., 2016. Ecosystem engineers in a self-organized soil: A review of concepts and future research questions. Soil Sci. 181, 91–109.

Lavonen, E.E., Kothawala, D.N., Tranvik, L.J., Gonsior, M., Schmitt-Kopplin, P., Köhler, S.J., 2015. Tracking changes in the optical properties and molecular composition of dissolved organic matter during drinking water production. Water Res. 85, 286–294. doi:10.1016/j.watres.2015.08.024

Lechtenfeld, O.J., Hertkorn, N., Shen, Y., Witt, M., Benner, R., 2015. Marine sequestration of carbon in bacterial metabolites. Nat. Commun. 6, 6711. doi:10.1038/ncomms7711

Lechtenfeld, O.J., Kattner, G., Flerus, R., McCallister, S.L., Schmitt-Kopplin, P., Koch, B.P., 2014. Molecular transformation and degradation of refractory dissolved organic matter in the Atlantic and Southern Ocean. Geochim. Cosmochim. Acta 126, 321–337. doi:10.1016/j.gca.2013.11.009

Ledesma, J.L.J., Grabs, T., Bishop, K.H., Schiff, S.L., Köhler, S.J., 2015. Potential for long-term transfer of dissolved organic carbon from riparian zones to streams in boreal catchments. Glob. Chang. Biol. 21, 2963–2979. doi:10.1111/gcb.12872

Lee, M.H., Park, J.-H., Matzner, E., 2018. Sustained production of dissolved organic carbon and nitrogen in forest floors during continuous leaching. Geoderma 310, 163–169. doi:10.1016/j.geoderma.2017.07.027

Leefmann, T., Frickenhaus, S., Koch, B.P., 2019. UltraMassExplorer: a browser-based application for the evaluation of high-resolution mass spectrometric data. Rapid Commun. Mass Spectrom. 33, 193–202. doi:10.1002/rcm.8315

Leenheer, J.A., 2009. Systematic approaches to comprehensive analyses of natural organic matter. Ann. Environ. Sci. 3, 1–130.

Leenheer, J.A., 1980. Origin and nature of humic substances in the waters of the Amazon River Basin. Acta Amaz. 10, 513–526. doi:10.1590/1809-43921980103513

Leenheer, J.A., Rostad, C.E., Gates, P.M., Furlong, E.T., Ferrer, I., 2001. Molecular resolution and fragmentation of fulvic acid by electrospray ionization/ multistage tandem mass spectrometry. Anal. Chem. 73, 1461–1471. doi:10.1021/ac0012593

Lehmann, J., Da Silva Cravo, M., Zech, W., 2001. Organic matter stabilization in a Xanthic Ferralsol of the central Amazon as affected by single trees: Chemical characterization of density, aggregate, and particle size fractions. Geoderma 99, 147–168. doi:10.1016/S0016-7061(00)00070-7

Lehmann, J., Kleber, M., 2015. The contentious nature of soil organic matter. Nature 528, 60–8. doi:10.1038/nature16069

Lehmann, K., Schaefer, S., Babin, D., Köhne, J.M., Schlüter, S., Smalla, K., Vogel, H.J., Totsche, K.U., 2018. Selective transport and retention of organic matter and bacteria shapes initial pedogenesis in artificial soil - A two-layer column study. Geoderma 325, 37–48. doi:10.1016/j.geoderma.2018.03.016

Leinemann, T., Preusser, S., Mikutta, R., Kalbitz, K., Cerli, C., Höschen, C., Mueller, C.W., Kandeler, E., Guggenberger, G., 2018. Multiple exchange processes on mineral surfaces control the transport of dissolved organic matter through soil profiles. Soil Biol. Biochem. 118, 79–90. doi:10.1016/j.soilbio.2017.12.006

Leopoldo, P.R., Matsui, E., Salati, E., Franken, W., Ribeiro, M. de N.G., 1982. Composição isotópica da água de chuva e da água do solo em floresta amazônica do tipo terra firme, região de Manaus. Acta Amaz. 12, 7–13. doi:10.1590/1809-43921982123s007

Leyva, D., Tose, L. V., Porter, J., Wolff, J., Jaffé, R., Fernandez-Lima, F., 2019. Understanding the structural complexity of dissolved organic matter: isomeric diversity. Faraday Discuss. Accepted A, Available Online.

doi:10.1039/c8fd00221e

Li, Y., Harir, M., Lucio, M., Gonsior, M., Koch, B.P., Schmitt-Kopplin, P., Hertkorn, N., 2016a. Comprehensive structure-selective characterization of dissolved organic matter by reducing molecular complexity and increasing analytical dimensions. Water Res. 106, 477–487. doi:10.1016/j.watres.2016.10.034

Li, Y., Harir, M., Lucio, M., Kanawati, B., Smirnov, K., Flerus, R., Koch, B.P., Schmitt-Kopplin, P., Hertkorn, N., 2016b. Proposed Guidelines for Solid Phase Extraction of Suwannee River Dissolved Organic Matter. Anal. Chem. 88, 6680–6688. doi:10.1021/acs.analchem.5b04501

Li, Y., Harir, M., Uhl, J., Kanawati, B., Lucio, M., Smirnov, K.S., Koch, B.P., Schmitt-Kopplin, P., Hertkorn, N., 2017. How representative are dissolved organic matter (DOM) extracts? A comprehensive study of sorbent selectivity for DOM isolation. Water Res. 116, 316–323.

Liang, C., Schimel, J.P., Jastrow, J.D., 2017. The importance of anabolism in microbial control over soil carbon storage. Nat. Microbiol. 2, 17105. doi:10.1038/nmicrobiol.2017.105

Linkhorst, A., Dittmar, T., Waska, H., 2016. Molecular Fractionation of Dissolved Organic Matter in a Shallow Subterranean Estuary: The Role of the Iron Curtain. Environ. Sci. Technol. 51, 1312–1320. doi:10.1021/acs.est.6b03608

Linz, A.M., He, S., Stevens, S.L.R., Anantharaman, K., Rohwer, R.R., Malmstrom, R.R., Bertilsson, S., McMahon, K.D., 2018. Freshwater carbon and nutrient cycles revealed through reconstructed population genomes. PeerJ 6, e6075. doi:10.7717/peerj.6075

Liu, X., Ser, Z., Cluntun, A.A., Mentch, S.J., Locasale, J.W., 2014. A Strategy for Sensitive, Large Scale Quantitative Metabolomics. J. Vis. Exp. 87. doi:10.3791/51358

Liu, Z., Sleighter, R.L., Zhong, J., Hatcher, P.G., 2011. The chemical changes of DOM from black waters to coastal marine waters by HPLC combined with ultrahigh resolution mass spectrometry. Estuar. Coast. Shelf Sci. 92, 205–216. doi:10.1016/j.ecss.2010.12.030

Longnecker, K., Kujawinski, E.B., 2016. Using network analysis to discern compositional patterns in ultrahigh-resolution mass spectrometry data of dissolved organic matter. Rapid Commun. Mass Spectrom. 30, 2388–2394. doi:10.1002/rcm.7719

Lu, K., Gardner, W.S., Liu, Z., 2018. Molecular structure characterization of riverine and coastal dissolved organic matter with ion mobility quadrupole time-of-flight LCMS (IM Q-TOF LCMS). Environ. Sci. Technol. 52, 7182–7191. doi:10.1021/acs.est.8b00999

Lucas, Y., 2001. The role of plants in controlling rates and products of weathering : Importance of biological pumping. Annu. Rev. Earth Planet. Sci. 29, 135–163. doi:10.1146/annurev.earth.29.1.135

Lucas, Y., Montes, C.R., Mounier, S., Loustau Cazalet, M., Ishida, D., Achard, R., Garnier, C., Coulomb, B., Melfi, A.J., 2012. Biogeochemistry of an Amazonian podzol-ferralsol soil system with white kaolin. Biogeosciences 9, 3705–3720. doi:10.5194/bg-9-3705-2012

Luek, J.L., Harir, M., Schmitt-Kopplin, P., Mouser, P.J., Gonsior, M., 2018. Organic sulfur fingerprint indicates continued injection fluid signature 10 months after hydraulic fracturing. Environ. Sci. Process. Impacts 21, 206–213. doi:10.1039/c8em00331a

Lutfalla, S., Barré, P., Bernard, S., Le Guillou, C., Alléon, J., Chenu, C., 2019. Multidecadal persistence of organic matter in soils: investigations at the submicrometer scale. Biogeosciences 16, 1401–1410. doi:10.5194/bg-2018-343

Lynch, L.M., Sutfin, N.A., Fegel, T.S., Boot, C.M., Covino, T.P., Wallenstein, M.D., 2019. River channel connectivity shifts metabolite composition and dissolved organic matter chemistry. Nat. Commun. 10, 459.

Ma, T., Zhu, S., Wang, Z., Chen, D., Dai, G., Feng, B., Su, X., Hu, H., Li, K., Han, W., Liang, C., Bai, Y., Feng, X., 2018. Divergent accumulation of microbial necromass and plant lignin components in grassland soils. Nat. Commun. 9, 3480. doi:10.1038/s41467-018-05891-1

Malik, A.A., Gleixner, G., 2013. Importance of microbial soil organic matter processing in dissolved organic carbon production. FEMS Microbiol. Ecol. 86, 139–148. doi:10.1111/1574-6941.12182

Malik, A.A., Roth, V.-N., Hébert, M., Tremblay, L., Dittmar, T., Gleixner, G., 2016. Linking molecular size, composition and carbon turnover of extractable soil microbial compounds. Soil Biol. Biochem. 100, 66–73. doi:10.1016/j.soilbio.2016.05.019

Mangal, V., Stock, N.L., Guéguen, C., 2016. Molecular characterization of phytoplankton dissolved organic matter (DOM) and sulfur components using high resolution Orbitrap mass spectrometry. Anal. Bioanal. Chem. 408, 1891–1900. doi:10.1007/s00216-015-9295-9

Mann, B.F., Chen, H., Herndon, E.M., Chu, R.K., Tolic, N., Portier, E.F., Chowdhury, T.R., Robinson, E.W., Callister, S.J., Wullschleger, S.D., Graham, D.E., Liang, L., Gu, B., 2015. Indexing permafrost soil organic matter degradation using high-resolution mass spectrometry. PLoS One 10, e0130557. doi:10.1371/journal.pone.0130557

Marín-Spiotta, E., Gruley, K.E., Crawford, J., Atkinson, E.E., Miesel, J.R., Greene, S., Cardona-Correa, C., Spencer, R.G.M., 2014. Paradigm shifts in soil organic matter research affect interpretations of aquatic carbon cycling: Transcending disciplinary and ecosystem boundaries. Biogeochemistry 117, 279–297. doi:10.1007/s10533-013-9949-7

Markewitz, D., Davidson, E.A., Figueiredo, R. de O., Victoria, R.L., Krusche, A. V., 2001. Control of cation concentrations in stream waters by surface soil processes in an Amazonian watershed. Nature 410, 802–805. doi:10.1038/35071052

Marques, J.D. de O., Libardi, P.L., Teixeira, W.G., Reis, A.M., 2004. Estudo de parâmetros físicos, químicos e hídricos de um Latossolo Amarelo, na região Amazônica. Acta Amaz. 34, 145–154. doi:10.1590/s0044-59672004000200002

Marques, J.D. de O., Teixeira, W.G., Reis, A.M., Cruz Junior, O.F., Batista, S.M., Afonso, M.A.C.B., 2010. Atributos químicos, físico-hídricos e mineralogia da fração argila em solos do Baixo Amazonas: Serra de Parintins. Acta Amaz. 40, 01–12. doi:10.1590/s0044-59672010000100001

Marschner, B., Brodowski, S., Dreves, A., Gleixner, G., Gude, A., Grootes, P.M., Hamer, U., Heim, A., Jandl, G., Ji, R., Kaiser, K., Kalbitz, K., Kramer, C., Leinweber, P., Rethemeyer, J., Schäffer, A., Schmidt, M.W.I., Schwark, L., Wiesenberg, G.L.B., 2008. How relevant is recalcitrance for the stabilization of organic matter in soils? J. Plant Nutr. Soil Sci. 171, 91–110. doi:10.1002/jpln.200700049

Marshall, A.G., Hendrickson, C.L., 2008. High-Resolution Mass Spectrometers. Annu. Rev. Anal. Chem. 1, 579–599. doi:10.1146/annurev.anchem.1.031207.112945

Marshall, A.G., Hendrickson, C.L., Jackson, G.S., 1998. Fourier transform ion cyclotron resonance mass spectrometry: A primer. Mass Spectrom. Rev. 17, 1–35. doi:10.1002/(SICI)1098-2787(1998)17:1<1::AID-MAS1>3.0.CO;2-K

Martin, J.C., Maillot, M., Mazerolles, G., Verdu, A., Lyan, B., Migné, C., Defoort, C., Canlet, C., Junot, C., Guillou, C., Manach, C., Jabob, D., Jouan-Rimbaud Bouveresse, D., Paris, E., Pujos-Guillot, E., Jourdan, F., Giacomoni, F., Courant, F., Favé, G., Le Gall, G., Chassaigne, H., Tabet, J.C., Martin, J.F., Antignac, J.P., Shintu, L., Defernez, M., Philo, M., Alexandre-Gouaubau, M.C., Amiot-Carlin, M.J., Bossis, M., Triba, M.N., Stojilkovic, N., Banzet, N., Molinié, R., Bott, R., Goulitquer, S., Caldarelli, S., Rutledge, D.N., 2015. Can we trust untargeted metabolomics? Results of the metabo-ring initiative, a large-scale, multi-instrument inter-laboratory study. Metabolomics 11, 807–821. doi:10.1007/s11306-014-0740-0

Bibliography

Masoom, H., Courtier-Murias, D., Farooq, H., Soong, R., Kelleher, B.P., Zhang, C., Maas, W.E., Fey, M., Kumar, R., Monette, M., Stronks, H.J., Simpson, M.J., Simpson, A.J., 2016. Soil Organic Matter in its native state: Unravelling the most complex biomaterial on earth. Environ. Sci. Technol. 50, 1670–1680. doi:10.1021/acs.est.5b03410

Mawhinney, D.B., Rosario-Ortiz, F., Baik, S., Vanderford, B.J., L.Snyder, S.A., 2009. Characterization of fulvic acids by liquid chromatography-quadrupole time-of-flight mass spectrometry. J. Chromatogr. A 1216, 1319–1324. doi:10.1016/j.chroma.2008.12.068

Mayorga, E., Aufdenkampe, A.K., Masiello, C. a, Krusche, A. V, Hedges, J.I., Quay, P.D., Richey, J.E., Brown, T.A., 2005. Young organic matter as a source of carbon dioxide outgassing from Amazonian rivers. Nature 436, 538–541. doi:10.1038/nature03880

McClain, M.E., Naiman, R.J., 2008. Andean influences on the biogeochemistry and ecology of the Amazon River. Bioscience 58, 325–338. doi:10.1641/B580408

McClain, M.E., Richey, J.E., Brandes, J.A., Pimentel, T.P., 1997. Dissolved organic matter and terrestrial-lotic linkages in the central Amazon basin of Brazil. Global Biogeochem. Cycles 11, 295–311. doi:10.1029/97GB01056

McIntyre, C., McRae, C., Jardine, D., Batts, B.D., 2002. Identification of compound classes in soil and peat fulvic acids as observed by electrospray ionization tandem mass spectrometry. Rapid Commun. Mass Spectrom. 16, 1604–1609. doi:10.1002/rcm.761

Meade, R.H., Rayol, J.M., da Conceicão, S.C., Natividade, J.R.G., 1991. Backwater Effects in the Amazon River Basin of Brazil. Env. Geol Water Sci.

Medeiros, P.M., Seidel, M., Niggemann, J., Spencer, R.G.M., Hernes, P.J., 2016. A novel molecular approach for tracing terrigenous dissolved organic matter into the deep ocean. Global Biogeochem. Cycles 30, 1–11. doi:10.1002/2015GB005320

Mentges, A., Feenders, C., Seibt, M., Blasius, B., Dittmar, T., 2017. Functional Molecular Diversity of Marine Dissolved Organic Matter Is Reduced during Degradation. Front. Mar. Sci. 4, 194. doi:10.3389/fmars.2017.00194

Michalzik, B., Kalbitz, K., Park, J.H., Solinger, S., Matzner, E., 2001. Fluxes and concentrations of dissolved organic carbon and nitrogen - A synthesis for temperate forests. Biogeochemistry 52, 173–205. doi:10.1023/A:1006441620810

Miguez-Macho, G., Fan, Y., 2012. The role of groundwater in the Amazon water cycle: 1. Influence on seasonal streamflow, flooding and wetlands. J. Geophys. Res. Atmos. 117, D15113. doi:10.1029/2012JD017539

Miketova, P., Schram, K.H., Whitney, J., Li, M., Huang, R., Kerns, E., Valcic, S., Timmermann, B.N., Rourick, R., Klohr, S., 2000. Tandem mass spectrometry studies of green tea catechins. Identification of three minor components in the polyphenolic extract of green tea. J. Mass Spectrom. 35, 860–869. doi:10.1002/1096-9888(200007)35:7<860::AID-JMS10>3.0.CO;2-J

Miltner, A., Bombach, P., Schmidt-Brücken, B., Kästner, M., 2012. SOM genesis: microbial biomass as a significant source. Biogeochemistry 111, 41–55. doi:10.1007/s10533-011-9658-z

Minor, E.C., Swenson, M.M., Mattson, B.M., Oyler, A.R., 2014. Structural characterization of dissolved organic matter: a review of current techniques for isolation and analysis. Environ. Sci. Process. Impacts 16, 2064–2079. doi:10.1039/C4EM00062E

Monteiro, M.T.F., Oliveira, S.M., Luizão, F.J., Cândido, L.A., Ishida, F.Y., Tomasella, J., 2014. Dissolved organic carbon concentration and its relationship to electrical conductivity in the waters of a stream in a forested Amazonian blackwater catchment. Plant Ecol. Divers. 7, 205–213. doi:10.1080/17550874.2013.820223

Montes, C.R., Lucas, Y., Pereira, O.J.R., Achard, R., Grimaldi, M., Melfi, A.J., 2011. Deep plant-derived carbon storage in Amazonian podzols. Biogeosciences 8, 113–120. doi:10.5194/bg-8-113-2011

Moritz, F., Kaling, M., Schnitzler, J.P., Schmitt-Kopplin, P., 2017. Characterization of poplar metabotypes via mass difference enrichment analysis. Plant Cell Environ. 40, 1057–1073. doi:10.1111/pce.12878

Mosher, J.J., Kaplan, L. a., Podgorski, D.C., McKenna, A.M., Marshall, A.G., 2015. Longitudinal shifts in dissolved organic matter chemogeography and chemodiversity within headwater streams: a river continuum reprise. Biogeochemistry 124, 371–385. doi:10.1007/s10533-015-0103-6

Mostovaya, A., Hawkes, J.A., Koehler, B., Dittmar, T., Tranvik, L.J., 2017. Emergence of the Reactivity Continuum of Organic Matter from Kinetics of a Multitude of Individual Molecular Constituents. Environ. Sci. Technol. 51, 11571–11579. doi:10.1021/acs.est.7b02876

Moyer, R.P., Bauer, J.E., Grottoli, A.G., 2013. Carbon isotope biogeochemistry of tropical small mountainous river, estuarine, and coastal systems of Puerto Rico. Biogeochemistry 112, 589–612. doi:10.1007/s10533-012-9751-y

Mullen, W., Yokota, T., Lean, M.E.J., Crozier, A., 2003. Analysis of ellagitannins and conjugates of ellagic acid and quercetin in raspberry fruits by LC-MSn. Phytochemistry 64, 617–624. doi:10.1016/S0031-9422(03)00281-4

Murphy, K.R., Timko, S.A., Gonsior, M., Powers, L.C., Wünsch, U.J., Stedmon, C.A., 2018. Photochemistry illuminates ubiquitous organic matter fluorescence spectra. Environ. Sci. Technol. 52, 11243–11250. doi:10.1021/acs.est.8b02648

Nagy, L., Artaxo, P., Forsberg, B.R., 2016. Interactions Between Biosphere, Atmosphere, and Human Land Use in the Amazon Basin: An introduction, in: Nagy, L., Artaxo, P., Forsberg, B.R. (Eds.), Interactions Between Biosphere, Atmosphere, and Human Land Use in the Amazon Basin. Ecological Studies 227. Springer, Heidelberg, pp. 3–15. doi:10.1007/978-3-662-49902-3_1

Ncube, E.N., Mhlongo, M.I., Piater, L.A., Steenkamp, P.A., Dubery, I.A., Madala, N.E., 2014. Analyses of chlorogenic acids and related cinnamic acid derivatives from Nicotiana tabacum tissues with the aid of UPLC-QTOF-MS/MS based on the in-source collision-induced dissociation method. Chem. Cent. J. 8, 1–10. doi:10.1186/s13065-014-0066-z

Nebbioso, A., Piccolo, A., 2015. Modification of chemical and conformational properties of natural organic matter by click chemistry as revealed by ESI-Orbitrap mass spectrometry. Anal. Bioanal. Chem. 407, 8515–8523. doi:10.1007/s00216-015-9005-7

Nebbioso, A., Piccolo, A., Spiteller, M., 2010. Limitations of electrospray ionization in the analysis of a heterogeneous mixture of naturally ocurring hydrophilic and hydrophobic compounds. Rapid Commun. Mass Spectrom. 24, 3163–3170. doi:10.1002/rcm

Newcomb, C.J., Qafoku, N.P., Grate, J.W., Bailey, V.L., De Yoreo, J.J., 2017. Developing a molecular picture of soil organic matter-mineral interactions by quantifying organo-mineral binding. Nat. Commun. 8, 396. doi:10.1038/s41467-017-00407-9

Nielsen, N.J., Christensen, P., Stedmon, C., Christensen, J.H., 2018. Examples of unwanted variation when characterising dissolved organic matter using direct injection electrospray mass spectrometry and chemometrics. Anal. Methods 10, 2636–2646. doi:10.1039/c8ay00226f

Nimmagadda, R.D., McRae, C., 2007. Characterisation of the backbone structures of several fulvic acids using a novel selective chemical reduction method. Org. Geochem. 38, 1061–1072. doi:10.1016/j.orggeochem.2007.02.016

Noacco, V., Duffy, C.J., Wagener, T., Worrall, F., Fasiolo, M., Howden, N.J.K., 2019. Drivers of interannual and

intra-annual variability of dissolved organic carbon concentration in the River Thames between 1884 and 2013. Hydrol. Process. 33, 994–1012. doi:10.1002/hyp.13379

Noriega-Ortega, B.E., Wienhausen, G., Mentges, A., Dittmar, T., Simon, M., Niggemann, J., 2019. Does the Chemodiversity of Bacterial Exometabolomes Sustain the Chemodiversity of Marine Dissolved Organic Matter? Front. Microbiol. 10, 215. doi:10.3389/fmicb.2019.00215

Novotny, N.R., Capley, E.N., Stenson, A.C., 2014. Fact or artifact: the representativeness of ESI-MS for complex natural organic mixtures. J. Mass Spectrom. 49, 316–26. doi:10.1002/jms.3345

Nowak, M.E., Schwab, V.F., Lazar, C.S., Behrendt, T., Kohlhepp, B., Totsche, K.U., Küsel, K., Trumbore, S.E., 2017. Carbon isotopes of dissolved inorganic carbon reflect utilization of different carbon sources by microbial communities in two limestone aquifer assemblages. Hydrol. Earth Syst. Sci. 21, 4283–4300. doi:10.5194/hess-21-4283-2017

Ohno, T., He, Z., Sleighter, R.L., Honeycutt, C.W., 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. doi:10.1021/es101089t

Osterholz, H., Niggemann, J., Giebel, H.-A., Simon, M., Dittmar, T., 2015. Inefficient microbial production of refractory dissolved organic matter in the ocean. Nat. Commun. 6, 7422. doi:10.1038/ncomms8422

Osterholz, H., Singer, G., Wemheuer, B., Daniel, R., Simon, M., Niggemann, J., Dittmar, T., 2016. Deciphering associations between dissolved organic molecules and bacterial communities in a pelagic marine system. ISME J. 10, 1717–1730. doi:10.1038/ismej.2015.231

Pacific, V.J., Jencso, K.G., McGlynn, B.L., 2010. Variable flushing mechanisms and landscape structure control stream DOC export during snowmelt in a set of nested catchments. Biogeochemistry 99, 193–211. doi:10.1007/s10533-009-9401-1

Patel-Sorrentino, N., Lucas, Y., Eyrolle, F., Melfi, A.J., 2007. Fe, Al and Si species and organic matter leached off a ferrallitic and podzolic soil system from Central Amazonia. Geoderma 137, 444–454. doi:10.1016/j.geoderma.2006.10.002

Perdue, E.M., Hertkorn, N., Kettrup, A., 2007. Substitution patterns in aromatic rings by increment analysis. Model development and application to natural organic matter. Anal. Chem. 79, 1010–1021. doi:10.1021/ac061611y

Perdue, E.M., Ritchie, J.D., 2014. Dissolved Organic Matter in freshwaters, in: Holland, H., Turekian, K., Drever, J.I. (Eds.), Treatise on Geochemistry: Second Edition. Volume 7: Surface and Groundwater, Weathering and Soils. Elsevier Ltd., Oxford, pp. 237–272. doi:10.1016/B978-0-08-095975-7.00201-1

Pereira, O.J.R., Montes, C.R., Lucas, Y., Santin, R.C., Melfi, A.J., 2015. A multi-sensor approach for mapping plant-derived carbon storage in Amazonian podzols. Int. J. Remote Sens. 36, 2076–2092. doi:10.1080/01431161.2015.1034896

Perry, R.H., Cooks, G., Noll, R.J., 2008. Orbitrap Mass spectrometry: Instrumentation, ion motion and applications. Mass Spectrom. Rev. 27, 661–699. doi:10.1002/mas.20186

Perveen, N., Barot, S., Maire, V., Cotrufo, M.F., Shahzad, T., Blagodatskaya, E., Stewart, C.E., Ding, W., Siddiq, M.R., Dimassi, B., Mary, B., Fontaine, S., 2019. Universality of priming effect: An analysis using thirty five soils with contrasted properties sampled from five continents. Soil Biol. Biochem. 134, 162–171. doi:10.1016/j.soilbio.2019.03.027

Petras, D., Koester, I., Da Silva, R., Stephens, B.M., Haas, A.F., Nelson, C.E., Kelly, L.W., Aluwihare, L.I., Dorrestein, P.C., 2017. High-resolution liquid chromatography tandem mass spectrometry enables large scale molecular characterization of dissolved organic matter. Front. Mar. Sci. 4, 406.

doi:10.3389/fmars.2017.00405

Peyton Smith, A., Bond-Lamberty, B., Benscoter, B.W., Tfaily, M.M., Hinkle, C.R., Liu, C., Bailey, V.L., 2017. Shifts in pore connectivity from precipitation versus groundwater rewetting increases soil carbon loss after drought. Nat. Commun. 8, 1335. doi:10.1038/s41467-017-01320-x

Peyton Smith, A., Marín-Spiotta, E., Balser, T., 2015. Successional and seasonal variations in soil and litter microbial community structure and function during tropical postagricultural forest regeneration: A multiyear study. Glob. Chang. Biol. 21, 3532–3547. doi:10.1111/gcb.12947

Peyton Smith, A., Marín-Spiotta, E., de Graaff, M.A., Balser, T.C., 2014. Microbial community structure varies across soil organic matter aggregate pools during tropical land cover change. Soil Biol. Biochem. 77, 292–303. doi:10.1016/j.soilbio.2014.05.030

Piccolo, A., 2001. The Supramolecular Structure of Humic Substances. Soil Sci. 166, 810–832. doi:10.1097/00010694-200111000-00007

Pitumpe-Arachchige, P.S., Hettiarachchi, G.M., Rice, C.W., Dynes, J.J., Maurmann, L., Wang, J., Karunakaran, C., Kilcoyne, A.L.D., Attanayake, C.P., Amado, T.J.C., Fiorin, J.E., 2018. Sub-micron level investigation reveals the inaccessibility of stabilized carbon in soil microaggregates. Sci. Rep. 8, 16810. doi:10.1038/s41598-018-34981-9

Pohlabeln, A.M., Dittmar, T., 2015. Novel insights into the molecular structure of non-volatile marine dissolved organic sulfur. Mar. Chem. 168, 86–94. doi:10.1016/j.marchem.2014.10.018

Pomerantz, A.E., Mullins, O.C., Paul, G., Ruzicka, J., Sanders, M., 2011. Orbitrap mass spectrometry: A proposal for routine analysis of nonvolatile components of petroleum. Energy and Fuels 25, 3077–3082. doi:10.1021/ef200359n

Poon, G.K., 1998. Analysis of catechins in tea extracts by liquid chromatography-electrospray ionization mass spectrometry. J. Chromatogr. A 794, 63–74. doi:10.1016/S0021-9673(97)01050-9

Preston, C.M., Nault, J.R., Trofymow, J.A., 2009. Chemical changes during 6 years of decomposition of 11 litters in some Canadian forest sites. Part 2. 13C abundance, solid-state 13C NMR spectroscopy and the meaning of "lignin." Ecosystems 12, 1078–1102. doi:10.1007/s10021-009-9267-z

Probandt, D., Eickhorst, T., Ellrott, A., Amann, R., Knittel, K., 2018. Microbial life on a sand grain: From bulk sediment to single grains. ISME J. 12, 623–633. doi:10.1038/ismej.2017.197

Qi, Y., Hempelmann, R., Volmer, D.A., 2016. Two-dimensional mass defect matrix plots for mapping genealogical links in mixtures of lignin depolymerisation products. Anal. Bioanal. Chem. 408, 4835–4843. doi:10.1007/s00216-016-9598-5

Qi, Y., O'Connor, P.B., 2014. Data Processing in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. Mass Spectrom. Rev. 33, 333–352.

Quan, Z., Huang, B., Lu, C., Shi, Y., Chen, X., Zhou, J., Fang, Y., 2018. Formation of extractable organic nitrogen in an agricultural soil: A 15N labeling study. Soil Biol. Biochem. 118, 161–165. doi:10.1016/j.soilbio.2017.12.015

Raeke, J., Lechtenfeld, O.J., Tittel, J., Oosterwoud, M.R., Bornmann, K., Reemtsma, T., 2017. Linking the mobilization of dissolved organic matter in catchments and its removal in drinking water treatment to its molecular characteristics. Water Res. 113, 149–159. doi:10.1016/j.watres.2017.01.066

Raeke, J., Lechtenfeld, O.J., Wagner, M., Herzsprung, P., Reemtsma, T., 2016. Selectivity of solid phase extraction of freshwater dissolved organic matter and its effect on ultrahigh resolution mass spectra. Environ. Sci. Process. Impacts 18, 918–927. doi:10.1039/C6EM00200E

Rasche, F., Scheubert, K., Hufsky, F., Zichner, T., Kai, M., Svatoš, A., Böcker, S., 2012. Identifying the unknowns by aligning fragmentation trees. Anal. Chem. 84, 3417–3426. doi:10.1021/ac300304u

Raymond, P.A., Saiers, J.E., 2010. Event controlled DOC export from forested watersheds. Biogeochemistry 100, 197–209. doi:10.1007/s10533-010-9416-7

Raymond, P.A., Spencer, R.G.M., 2014. Riverine DOM, in: Hansell, D.A., Carlson, C.A. (Eds.), Biogeochemistry of Marine Dissolved Organic Matter: Second Edition. Academic Press, Cambridge, pp. 509–533. doi:10.1016/B978-0-12-405940-5.00011-X

Reemtsma, T., 2010. The carbon versus mass diagram to visualize and exploit FTICR-MS data of natural organic matter. J. Mass Spectrom. 45, 382–390. doi:10.1002/jms.1722

Remington, S.M., Strahm, B.D., Neu, V., Richey, J.E., Da Cunha, H.B., 2007. The role of sorption in control of riverine dissolved organic carbon concentrations by riparian zone soils in the Amazon basin. Soil Sci. 172, 279–291. doi:10.1097/ss.0b013e318032ab46

Remucal, C.K., Cory, R.M., Sander, M., McNeill, K., 2012. Low molecular weight components in an aquatic humic substance as characterized by membrane dialysis and orbitrap mass spectrometry. Environ. Sci. Technol. 46, 9350–9. doi:10.1021/es302468q

Repeta, D.J., 2015. Chemical characterization and cycling of dissolved organic matter, in: Hansell, D.A., Carlson, C.A. (Eds.), Biogeochemistry of Marine Dissolved Organic Matter. Elsevier Inc., London, pp. 21–63. doi:10.1016/B978-0-12-405940-5.00002-9

Richardson, J.B., 2017. Critcial Zone, in: White, W. (Ed.), Encyclopedia of Geochemistry. Encyclopedia of Earth Sciences Series. Springer, Cham. doi:10.2138/gselements.15.2.137a

Riedel, T., Dittmar, T., 2014. A Method Detection Limit for the Analysis of Natural Organic Matter via Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. Anal. Chem. 86, 8376–82. doi:10.1021/ac501946m

Riedel, T., Zak, D., Biester, H., Dittmar, T., 2013. Iron traps terrestrially derived dissolved organic matter at redox interfaces. PNAS 110, 10101–5. doi:10.1073/pnas.1221487110

Riedel, T., Zark, M., Vähätalo, A. V., Niggemann, J., Spencer, R.G.M., Hernes, P.J., Dittmar, T., 2016. Molecular Signatures of Biogeochemical Transformations in Dissolved Organic Matter from Ten World Rivers. Front. Earth Sci. 4, 85. doi:10.3389/feart.2016.00085

Rockenbach, I.I., Jungfer, E., Ritter, C., Santiago-Schübel, B., Thiele, B., Fett, R., Galensa, R., 2012. Characterization of flavan-3-ols in seeds of grape pomace by CE, HPLC-DAD-MS n and LC-ESI-FTICR-MS. Food Res. Int. 48, 848–855. doi:10.1016/j.foodres.2012.07.001

Rodgers, R.P., Mapolelo, M.M., Robbins, W.K., Chacón-Patiño, M.L., Putman, J.C., Niles, S.F., Rowland, S.M., Marshall, A.G., 2019. Combating selective ionization in the high resolution mass spectral characterization of complex mixtures. Faraday Discuss. Available online. doi:10.1039/C9FD00005D

Rogers, S., Wei Ong, C., Wandy, J., Ernst, M., Ridder, L., van der Hooft, J.J.J., 2019. Deciphering complex metabolite mixtures by unsupervised and supervised substructure discovery and semi-automated annotation from MS/MS spectra. Faraday Discuss. Accepted M, Available Online. doi:10.1039/c8fd00235e

Rossel, P.E., Bienhold, C., Boetius, A., Dittmar, T., 2016. Dissolved organic matter in pore water of Arctic Ocean sediments: Environmental influence on molecular composition. Org. Geochem. 97, 41–52. doi:10.1016/j.orggeochem.2016.04.003

Rossel, P.E., Vähätalo, A. V., Witt, M., Dittmar, T., 2013. Molecular composition of dissolved organic matter from a wetland plant (Juncus effusus) after photochemical and microbial decomposition (1.25 yr): Common

features with deep sea dissolved organic matter. Org. Geochem. 60, 62–71. doi:10.1016/j.orggeochem.2013.04.013

Roth, V.-N., Dittmar, T., Gaupp, R., Gleixner, G., 2015. The Molecular Composition of Dissolved Organic Matter in Forest Soils as a Function of pH and Temperature. PLoS One 10, e0119188. doi:10.1371/journal.pone.0119188

Roth, V.-N., Dittmar, T., Gaupp, R., Gleixner, G., 2014. Ecosystem-specific composition of dissolved organic matter. Vadose Zo. J. 13. doi:http://dx.doi.org/10.2136/vzj2013.09.0162

Roth, V.-N., Dittmar, T., Gaupp, R., Gleixner, G., 2013. Latitude and pH driven trends in the molecular composition of DOM across a north south transect along the Yenisei River. Geochim. Cosmochim. Acta 123, 93–105. doi:10.1016/j.gca.2013.09.002

Roth, V.-N., Lange, M., Simon, C., Hertkorn, N., Bucher, S., Goodall, T., Griffiths, R.I., Mellado-Vázquez, P.G., Mommer, L., Oram, N.J., Weigelt, A., Dittmar, T., Gleixner, G., accepted. Molecular and structural changes of dissolved organic matter explain its persistence in the Critical Zone. Nat. Geosci.

Rousk, J., Bååth, E., Brookes, P.C., Lauber, C.L., Lozupone, C., Caporaso, J.G., Knight, R., Fierer, N., 2010. Soil bacterial and fungal communities across a pH gradient in an arable soil. ISME J. 4, 1340–1351. doi:10.1038/ismej.2010.58

Sachse, D., Kahmen, A., Gleixner, G., 2009. Significant seasonal variation in the hydrogen isotopic composition of leaf-wax lipids for two deciduous tree ecosystems (Fagus sylvativa and Acer pseudoplatanus). Org. Geochem. 40, 732–742. doi:10.1016/j.orggeochem.2009.02.008

Saldanha, L.L., Vilegas, W., Dokkedal, A.L., 2013. Characterization of flavonoids and phenolic acids in Myrcia bella Cambess. Using FIA-ESI-IT-MSn and HPLC-PAD-ESI-IT-MS combined with NMR. Molecules 18, 8402–8416. doi:10.3390/molecules18078402

Sanderman, J., Hengl, T., Fiske, G.J., 2017. Soil carbon debt of 12,000 years of human land use. Proc. Natl. Acad. Sci. 114, 9575–9580. doi:10.1073/pnas.1800925115

Sanderman, J., Lohse, K.A., Baldock, J.A., Amundson, R., 2009. Linking soils and streams: Sources and chemistry of dissolved organic matter in a small coastal watershed. Water Resour. Res. 45, W03418. doi:10.1029/2008WR006977

Sandron, S., Rojas, A., Wilson, R., Davies, N.W., Haddad, P.R., Shellie, R.A., Nesterenko, P.N., Kelleher, B.P., Paull, B., 2015. Chromatographic methods for the isolation, separation and characterisation of dissolved organic matter. Environ. Sci. Process. Impacts 17, 1531–1567. doi:10.1039/c5em00223k

Šantl-Temkiv, T., Finster, K., Dittmar, T., Hansen, B.M., Thyrhaug, R., Nielsen, N.W., Karlson, U.G., 2013. Hailstones: A Window into the Microbial and Chemical Inventory of a Storm Cloud. PLoS One 8. doi:10.1371/journal.pone.0053550

Saunders, T.J., McClain, M.E., Llerena, C.A., 2006. The biogeochemistry of dissolved nitrogen, phosphorus, and organic carbon along terrestrial-aquatic flowpaths of a montane headwater catchment in the Peruvian Amazon. Hydrol. Process. 20, 2549–2562. doi:10.1002/hyp

Scheibe, A., Gleixner, G., 2014. Influence of Litter Diversity on Dissolved Organic Matter Release and Soil Carbon Formation in a Mixed Beech Forest. PLoS One 9, e114040. doi:10.1371/journal.pone.0114040

Scheibe, A., Steffens, C., Seven, J., Jacob, A., Hertel, D., Leuschner, C., Gleixner, G., 2015. Effects of tree identity dominate over tree diversity on the soil microbial community structure. Soil Biol. Biochem. 81, 219–227. doi:10.1016/j.soilbio.2014.11.020

Schimel, J.P., 2018. Life in dry soils: Effects of drought on soil microbial communities and processes. Annu. Rev.

Ecol. Evol. Syst. 49, 409–432. doi:10.1146/annurev-ecolsys-110617-062614

Schimel, J.P., Schaeffer, S.M., 2012. Microbial control over carbon cycling in soil. Front. Microbiol. 3, 348. doi:10.3389/fmicb.2012.00348

Schlüter, S., Eickhorst, T., Mueller, C.W., 2019. Correlative imaging reveals holistic view of soil microenvironments. Environ. Sci. Technol. 53, 829–837. doi:10.1021/acs.est.8b05245

Schmidt, M.W.I., Torn, M.S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I.A., Kleber, M., Kögel-Knabner, I., Lehmann, J., Manning, D.A.C., Nannipieri, P., Rasse, D.P., Weiner, S., Trumbore, S.E., 2011. Persistence of soil organic matter as an ecosystem property. Nature 478, 49–56. doi:10.1038/nature10386

Schulze, W.X., Gleixner, G., Kaiser, K., Guggenberger, G., Mann, M., Schulze, E.D., 2005. A proteomic fingerprint of dissolved organic carbon and of soil particles. Oecologia 142, 335–343. doi:10.1007/s00442-004-1698-9

Schwab, V.F., Hermann, M., Roth, V.-N., Gleixner, G., Lehmann, R., Pohnert, G., Trumbore, S., Küsel, K., Totsche, K.U., 2017. Functional diversity of microbial communities in pristine aquifers inferred by PLFA- and sequencing-based approaches. Biogeosciences 14, 2697–2714. doi:10.5194/bg-2016-442

Seibert, J., Grabs, T., Köhler, S., Laudon, H., Winterdahl, M., Bishop, K., 2009. Linking soil- and stream-water chemistry based on a riparian flow-concentration integration model. Hydrol. Earth Syst. Sci. 13, 2287–2297. doi:10.5194/hess-13-2287-2009

Seidel, M., Beck, M., Riedel, T., Waska, H., Suryaputra, I.G.N.A., Schnetger, B., Niggemann, J., Simon, M., Dittmar, T., 2014. Biogeochemistry of dissolved organic matter in an anoxic intertidal creek bank. Geochim. Cosmochim. Acta 140, 418–434. doi:10.1016/j.gca.2014.05.038

Seifert, A.-G., Roth, V.-N., Dittmar, T., Gleixner, G., Breuer, L., Houska, T., Marxsen, J., 2016. Comparing molecular composition of dissolved organic matter in soil and stream water: Influence of land use and chemical characteristics. Sci. Total Environ. 571, 142–152. doi:10.1016/j.scitotenv.2016.07.033

Seyler, F., Muller, F., Cochonneau, G., Guimarães, L., Guyot, J.L., 2009. Watershed delineation for the Amazon sub-basin system using GTOPO30 DEM and a drainage network extracted from JERS SAR images. Hydrol. Process. 23, 3173–3185. doi:10.1002/hyp

Sierra, C.A., Hoyt, A.M., He, Y., Trumbore, S.E., 2018. Soil organic matter persistence as a stochastic process: age and transit time distributions of carbon in soils. Global Biogeochem. Cycles 32, 1574–1588. doi:10.1029/2018GB005950

Silvério, D. V., Brando, P.M., Bustamante, M.M.C., Putz, F.E., Marra, D.M., Levick, S.R., Trumbore, S.E., 2019. Fire, fragmentation, and windstorms: A recipe for tropical forest degradation. J. Ecol. 107, 656–667. doi:10.1111/1365-2745.13076

Simon, C., Osterholz, H., Koschinsky, A., Dittmar, T., 2019. Riverine mixing at the molecular scale – An ultrahigh-resolution mass spectrometry study on dissolved organic matter and selected metals in the Amazon confluence zone (Manaus, Brazil). Org. Geochem. 129, 45–62. doi:10.1016/j.orggeochem.2019.01.013

Simon, C., Roth, V.-N., Dittmar, T., Gleixner, G., 2018. Molecular Signals of Heterogeneous Terrestrial Environments Identified in Dissolved Organic Matter: A Comparative Analysis of Orbitrap and Ion Cyclotron Resonance Mass Spectrometers. Front. Earth Sci. 6, 1–16. doi:10.3389/feart.2018.00138

Simpson, A.J., Simpson, M.J., Smith, E., Kelleher, B.P., 2007. Microbially derived inputs to soil organic matter: Are current estimates too low? Environ. Sci. Technol. 41, 8070–8076. doi:10.1021/es071217x

Sioli, H., 1954. Gewässerchemie und Vorgänge in den Böden im Amazonasgebiet. Naturwissenschaften 41, 456–457.

Six, J., Bossuyt, H., Degryze, S., Denef, K., 2004. A history of research on the link between (micro)aggregates, soil biota, and soil organic matter dynamics. Soil Tillage Res. 79, 7–31. doi:10.1016/j.still.2004.03.008

Sleighter, R.L., Hatcher, P.G., 2011. Fourier Transform mass spectrometry for the molecular level characterization of natural organic matter: Instrument capabilities, applications, and limitations, in: Nikolic, G. (Ed.), Fourier Transforms - Approach to Scientific Principles. InTech Europe, Rijeka, pp. 295–320. doi:10.5772/15959

Sleighter, R.L., McKee, G.A., Hatcher, P.G., 2009. Direct Fourier transform mass spectral analysis of natural waters with low dissolved organic matter. Org. Geochem. 40, 119–125. doi:10.1016/j.orggeochem.2008.09.012

Sleighter, R.L., McKee, G.A., Liu, Z., 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. doi:10.4319/lom.2008.6.246

Sleno, L., 2012. The use of mass defect in modern mass spectrometry. J. Mass Spectrom. 47, 226–236. doi:10.1002/jms.2953

Smirnov, K.S., Forcisi, S., Moritz, F., Lucio, M., Schmitt-Kopplin, P., 2019. Mass difference maps and their application for the re-calibration of mass spectrometric data in non-targeted metabolomics. Anal. Chem. doi:10.1021/acs.analchem.8b04555

Smith, D.F., Podgorski, D.C., Rodgers, R.P., Blakney, G.T., Hendrickson, C.L., 2018. 21 Tesla FT-ICR mass spectrometer for ultrahigh-resolution analysis of complex organic mixtures. Anal. Chem. 90, 2041–2047. doi:10.1021/acs.analchem.7b04159

Smith, E.A., Park, S., Klein, A.T., Lee, Y.J., 2012. Bio-Oil Analysis Using Negative Electrospray Ionization: Comparative Study of High Resolution Mass Spectrometers and Phenolic vs. Sugaric Components. Energy & Fuels 26, 3796–3802. doi:10.1021/ef3003558

Snyder, D.T., Fedick, P.W., Cooks, R.G., 2016. Multigenerational collision-induced dissociation for characterization of organic compounds. Anal. Chem. 88, 9572–9581. doi:10.1021/acs.analchem.6b02209

Sokol, N.W., Sanderman, J., Bradford, M.A., 2019. Pathways of mineral-associated soil organic matter formation: Integrating the role of plant carbon source, chemistry, and point of entry. Glob. Chang. Biol. 25, 12–24. doi:10.1111/gcb.14482

Sollins, P., Kramer, M.G., Swanston, C., Lajtha, K., Filley, T., Aufdenkampe, A.K., Wagai, R., Bowden, R.D., 2009. Sequential density fractionation across soils of contrasting mineralogy: Evidence for both microbial- and mineral-controlled soil organic matter stabilization. Biogeochemistry 96, 209–231. doi:10.1007/s10533-009-9359-z

Soucémarianadin, L.N., Erhagen, B., Nilsson, M.B., Öquist, M.G., Immerzeel, P., Schleucher, J., 2017. Two dimensional NMR spectroscopy for molecular characterization of soil organic matter: Application to boreal soils and litter. Org. Geochem. 113, 184–195. doi:10.1016/j.orggeochem.2017.06.019

Steinhof, A., Altenburg, M., Machts, H., 2017. Sample preparation at the Jena 14C laboratory. Radiocarbon 59, 815–830. doi:10.1017/RDC.2017.50

Steinhof, A., Hejja, I., Wagner, T., 2010. Improvements of the Jena AMS system. Nucl. Instruments Methods Phys. Res. Sect. B 268, 902–905. doi:10.1016/j.nimb.2009.10.060

Stenson, A.C., Marshall, A.G., Cooper, W.T., 2003. Exact Masses and Chemical Formulas of Individual Suwannee River Fulvic Acids from Ultrahigh Resolution Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectra Molecular formulas have been assigned for 4626 indi- mass measurements fr. Anal. Chem. 75, 1275–1284. doi:10.1021/ac026106p

Stenson, A.C., Ruddy, B.M., Bythell, B.J., 2014. International Journal of Mass Spectrometry Ion molecule reaction H / D exchange as a probe for isomeric fractionation in chromatographically separated natural organic matter. Int. J. Mass Spectrom. 360, 45–53. doi:10.1016/j.ijms.2013.12.026

Strobel, B.W., 2001. Influence of vegetation on low-molecular-weight carboxylic acids in soil solution - A review. Geoderma 99, 169–198. doi:10.1016/S0016-7061(00)00102-6

Strohalm, M., Kavan, D., Novák, P., Volný, M., Havlíček, V., 2010. MMass 3: A cross-platform software environment for precise analysis of mass spectrometric data. Anal. Chem. 82, 4648–4651. doi:10.1021/ac100818g

Stubbins, A., Spencer, R.G.M., Chen, H., Hatcher, P.G., Mopper, K., Hernes, P.J., Mwamba, V.L., Mangangu, A.M., Wabakanghanzi, J.N., 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. doi:10.4319/lo.2010.55.4.1467

Sun, S., Schefuß, E., Mulitza, S., Chiessi, C.M., Sawakuchi, A.O., Zabel, M., Baker, P.A., Hefter, J., Mollenhauer, G., 2017. Origin and processing of terrestrial organic carbon in the Amazon system: Lignin phenols in river, shelf, and fan sediments. Biogeosciences 14, 2495–2512. doi:10.5194/bg-14-2495-2017

Swenson, M.M., Oyler, A.R., Minor, E.C., 2014. Rapid solid phase extraction of dissolved organic matter. Limnol. Oceanogr. Methods 12, 713–728. doi:10.4319/lom.2014.12.713

Swenson, T.L., Jenkins, S., Bowen, B.P., Northen, T.R., 2015. Untargeted soil metabolomics methods for analysis of extractable organic matter. Soil Biol. Biochem. 80, 189–198. doi:10.1016/j.soilbio.2014.10.007

Targhetta, N., Kesselmeier, J., Wittmann, F., 2015. Effects of the hydroedaphic gradient on tree species composition and aboveground wood biomass of oligotrophic forest ecosystems in the central Amazon basin. Folia Geobot. 50, 185–205. doi:10.1007/s12224-015-9225-9

Tefs, C., Gleixner, G., 2012. Importance of root derived carbon for soil organic matter storage in a temperate old-growth beech forest - Evidence from C, N and 14C content. For. Ecol. Manage. 263, 131–137. doi:10.1016/j.foreco.2011.09.010

Tfaily, M.M., Chu, R.K., Tolić, N., Roscioli, K.M., Anderton, C.R., Paša-Tolić, L., Robinson, E.W., Hess, N.J., 2015. Advanced solvent based methods for molecular characterization of soil organic matter by high-resolution mass spectrometry. Anal. Chem. 87, 5206–5215. doi:10.1021/acs.analchem.5b00116

Thermo Fisher Scientific, 2013. Orbitrap Elite Hardware Manual, Revision A. Thermo Fisher Scientific, Bremen.

These, A., Winkler, M., Thomas, C., Reemtsma, T., 2004. Determination of molecular formulas and structural regularities of low molecular weight fulvic acids by size-exclusion chromatography with electrospray ionization quadrupole time-of-flight mass spectrometry. Rapid Commun. Mass Spectrom. 18, 1777–1786. doi:10.1002/rcm.1550

Thurman, E.M., 1985. Organic Geochemistry of Natural Waters. Martinus Nijhoff/ Dr W. Junk Publishers, Dordrecht. doi:10.1007/978-94-009-5095-5

Tomasella, J., Hodnett, M.G., Cuartas, L.A., Nobre, A.D., Waterloo, M.J., Oliveira, S.M., 2007. The water balance of an Amazonian micro-catchment: The effect of interannual variability of rainfall on hydrological behaviour. Hydrol. Process. 22, 2133–2147. doi:10.1002/hyp.6813

Torres-Sallan, G., Schulte, R.P.O., Lanigan, G.J., Byrne, K.A., Reidy, B., Simó, I., Six, J., Creamer, R.E., 2017. Clay illuviation provides a long-term sink for C sequestration in subsoils. Sci. Rep. 7, 45635. doi:10.1038/srep45635

Totsche, K.U., Amelung, W., Gerzabek, M.H., Guggenberger, G., Klumpp, E., Knief, C., Lehndorff, E., Mikutta,

R., Peth, S., Prechtel, A., Ray, N., Kögel-Knabner, I., 2018. Microaggregates in soils. J. Plant Nutr. Soil Sci. 181, 104–136. doi:10.1002/jpln.201600451

Townsend, A.R., Cleveland, C.C., Houlton, B.Z., Alden, C.B., White, J.W.C., 2011. Multi-element regulation of the tropical forest carbon cycle. Front. Ecol. Environ. 9, 9–17. doi:10.1890/100047

Trumbore, S.E., Sierra, C.A., Hicks Pries, C.E., 2016. Radiocarbon nomenclature, theory, models, and interpretation: measuring age, determining cycling rates, and tracing source pools, in: Schuur, E.A.G., Druffel, E.R.M., Trumbore, Susan E. (Eds.), Climate, Radiocarbon and Change. Springer, Basel, pp. 45–82.

Tunaley, C., Tetzlaff, D., Lessels, J., Soulsby, C., 2016. Linking high-frequency DOC dynamics to the age of connected water sources. Water Resour. Res. 52, 5232–5247. doi:10.1002/2015WR018419

Vidal, A., Watteau, F., Remusat, L., Mueller, C.W., Nguyen Tu, T.-T., Buegger, F., Derenne, S., Quenea, K., 2019. Earthworm cast formation and development: A shift from plant litter to mineral associated organic matter. Front. Environ. Sci. 7, 55. doi:10.3389/fenvs.2019.00055

Vogel, C., Mueller, C., Höschen, C., Buegger, F., Heister, K., Schulz, S., Schloter, M., Kögel-Knabner, I., 2014. Submicron structures provide preferential spots for carbon and nitrogen sequestration in soils. Nat. Commun. 5, 1–7.

Waggoner, D.C., Chen, H., Willoughby, A.S., 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. doi:10.1016/j.orggeochem.2015.02.007

Waggoner, D.C., Hatcher, P.G., 2017. Hydroxyl radical alteration of HPLC fractionated lignin: Formation of new compounds from terrestrial organic matter. Org. Geochem. doi:10.1016/j.orggeochem.2017.07.011

Waggoner, D.C., Wozniak, A.S., Cory, R.M., Hatcher, P.G., 2017. The role of reactive oxygen species in the degradation of lignin derived dissolved organic matter. Geochim. Cosmochim. Acta 208, 171–184. doi:10.1016/j.gca.2017.03.036

Wagner, S., Dittmar, T., Jaffé, R., 2015. Molecular characterization of dissolved black nitrogen via electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. Org. Geochem. 79, 21–30. doi:10.1016/j.orggeochem.2014.12.002

Wang, M., Carver, J.J., Phelan, V. V., Sanchez, L.M., Garg, N., Peng, Y., Nguyen, D.D., Watrous, J., Kapono, C.A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V., Meehan, M.J., [...], Gerwick, W.H., Moore, B.S., Dorrestein, P.C., Bandeira, N., 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat. Biotechnol. 34, 828–837. doi:10.1038/nbt.3597

Ward, N., Bianchi, T., Medeiros, P., Seidel, M., Richey, J., Keil, R., Sawakuchi, H., 2017. Where Carbon Goes When Water Flows: Carbon Cycling Across the Aquatic Continuum. Front. Ecol. Environ. 4, 1–27. doi:10.3389/fmars.2017.00007

Ward, N.D., Keil, R.G., Medeiros, P.M., Brito, D.C., Cunha, A.C., Dittmar, T., Yager, P.L., Krusche, A. V., Richey, J.E., 2013. Degradation of terrestrially derived macromolecules in the Amazon River. Nat. Geosci. 6, 530–533. doi:10.1038/ngeo1817

Waterloo, M.J., Oliveira, S.M., Drucker, D.P., Nobre, A.D., Cuartas, L.A., Hodnett, M.G., Langedijk, I., Jans, W.W.P., Tomasella, J., Araujo, A.C. de, Pimentel, T.P., Estrada, J.C.M., 2006. Export of organic carbon in run-off from an Amazonian rainforest blackwater catchment. Hydrol. Process. 20, 2581–2597.

Webb, J.R., Santos, I.R., Maher, D.T., Finlay, K., 2018. The importance of aquatic carbon fluxes in net ecosystem carbon budgets: A catchment-scale review. Ecosystems 22, 508–527.

Wells, M.J.M., Stretz, H.A., 2019. Supramolecular architectures of natural organic matter. Sci. Total Environ. 671, 1125–1133. doi:10.1016/j.scitotenv.2019.03.406

Witt, M., Fuchser, J., Koch, B.P., 2009. Fragmentation studies of fulvic acids using collision induced dissociation fourier transform ion cyclotron resonance mass spectrometry. Anal. Chem. 81, 2688–94. doi:10.1021/ac802624s

Woche, S.K., Goebel, M.O., Mikutta, R., Schurig, C., Kaestner, M., Guggenberger, G., Bachmann, J., 2017. Soil wettability can be explained by the chemical composition of particle interfaces - An XPS study. Sci. Rep. 7, 42877. doi:10.1038/srep42877

Wolfender, J.-L., Nuzillard, J.-M., Van Der Hooft, J.J.J., Renault, J.-H., Bertrand, S., 2019. Accelerating metabolite identification in natural product research: Toward an ideal combination of liquid chromatography-high-resolution tandem mass spectrometry and NMR profiling, in silico databases, and chemometrics. Anal. Chem. 91, 704–742. doi:10.1021/acs.analchem.8b05112

Woods, G.C., Simpson, M.J., Simpson, A.J., 2012. Oxidized sterols as a significant component of dissolved organic matter: Evidence from 2D HPLC in combination with 2D and 3D NMR spectroscopy. Water Res. 46, 3398–3408. doi:10.1016/j.watres.2012.03.040

Woolf, D., Lehmann, J., 2019. Microbial models with minimal mineral protection can explain long-term soil organic carbon persistence. Sci. Rep. 9, 6522. doi:10.1038/s41598-019-43026-8

Wynn, J.G., 2007. Carbon isotope fractionation during decomposition of organic matter in soils and paleosols: Implications for paleoecological interpretations of paleosols. Palaeogeogr. Palaeoclimatol. Palaeoecol. 251, 437–448. doi:10.1016/j.palaeo.2007.04.009

Wyrepkowski, C.C., Da Costa, D.L.M.G., Sinhorin, A.P., Vilegas, W., De Grandis, R.A., Resende, F.A., Varanda, E.A., Dos Santos, L.C., 2014. Characterization and quantification of the compounds of the ethanolic extract from Caesalpinia ferrea stem bark and evaluation of their mutagenic activity. Molecules 19, 16039–16057. doi:10.3390/molecules191016039

Yuzuak, S., Ballington, J., Xie, D.-Y., 2018. HPLC-qTOF-MS/MS-Based Profiling of Flavan-3-ols and Dimeric Proanthocyanidins in Berries of Two Muscadine Grape Hybrids FLH 13-11 and FLH 17-66. Metabolites 8, 57. doi:10.3390/metabo8040057

Zanchi, F.B., Meesters, A.G.C.A., Waterloo, M.J., Kruijt, B., Kesselmeier, J., Luizão, F.J., Dolman, A.J., 2014. Soil CO2 exchange in seven pristine Amazonian rain forest sites in relation to soil temperature. Agric. For. Meteorol. 192–193, 96–107. doi:10.1016/j.agrformet.2014.03.009

Zanchi, F.B., Waterloo, M.J., Dolman, A.J., Groenendijk, M., Kesselmeier, J., Kruijt, B., Bolson, M.A., Luizão, F.J., Manzi, A.O., 2011. Influence of drainage status on soil and water chemistry, litter decomposition and soil respiration in central Amazonian forests on sandy soils. Rev. Ambient. e Agua 6, 6–29. doi:10.4136/1980-993X

Zanchi, F.B., Waterloo, M.J., Tapia, A.P., Alvarado Barrientos, M.S., Bolson, M.A., Luizão, F.J., Manzi, A.O., Dolman, A.J., 2015. Water balance, nutrient and carbon export from a heath forest catchment in central Amazonia, Brazil. Hydrol. Process. 29, 3633–3648. doi:10.1002/hyp.10458

Zark, M., Christoffers, J., Dittmar, T., 2017. Molecular properties of deep-sea dissolved organic matter are predictable by the central limit theorem: Evidence from tandem FT-ICR-MS. Mar. Chem. 191, 9–15. doi:10.1016/j.marchem.2017.02.005

Zark, M., Dittmar, T., 2018. Universal molecular structures in natural dissolved organic matter. Nat. Commun. 9, 3178. doi:10.1038/s41467-018-05665-9

Zarnetske, J.P., Bouda, M., Abbott, B.W., Saiers, J., Raymond, P.A., 2018. Generality of hydrologic transport

limitation of watershed organic Carbon flux across ecoregions of the United States. Geophys. Res. Lett. 45, 11702–11711. doi:10.1029/2018GL080005

Zavarzina, A.G., Lisov, A.A., Zavarzin, A.A., Leontievsky, A.A., 2011. Fungal oxidoreductases and humification in forest soils, in: Shukla, G., Varma, A. (Eds.), Soil Enzymology. Springer, Heidelberg, pp. 207–228.

Zavarzina, A.G., Lisov, A. V., Leontievsky, A.A., 2018. The role of ligninolytic enzymes laccase and a versatile peroxidase of the white-rot fungus Lentinus tigrinus in biotransformation of soil humic matter: Comparative in vivo study. J. Geophys. Res. Biogeosciences 123, 2727–2742. doi:10.1029/2017JG004309

Zhang, F., Harir, M., Moritz, F., Zhang, J., Witting, M., Wu, Y., Schmitt-Kopplin, P., Fekete, A., Gaspar, A., Hertkorn, N., 2014. Molecular and structural characterization of dissolved organic matter during and post cyanobacterial bloom in Taihu by combination of NMR spectroscopy and FTICR mass spectrometry. Water Res. 57C, 280–294. doi:10.1016/j.watres.2014.02.051

Zhang, X.-P., Yang, Z.-L., Niu, G.-Y., Wang, X.-Y., 2009. Stable water isotope simulation in different reservoirs of Manaus, Brazil, by Community Land Model incorporating stable isotopic effect. Int. J. Climatol. 29, 619–628. doi:10.1002/joc

Zherebker, A.Y., Airapetyan, D., Konstantinov, A.I., Kostyukevich, Y.I., Kononikhin, A.S., Popov, I.A., Zaitsev, K. V., Nikolaev, E.N., Perminova, I. V., 2015. Synthesis of model humic substances: A mechanistic study using controllable H/D exchange and Fourier transform ion cyclotron resonance mass spectrometry. Analyst 140, 4708–4719. doi:10.1039/c5an00602c

Zhurov, K.O., Kozhinov, A.N., Tsybin, Y.O., 2013. Evaluation of high-field orbitrap fourier transform mass spectrometer for petroleomics. Energy & Fuels 27, 2974–2983. doi:10.1021/ef400203g

Zsolnay, Á., 2003. Dissolved organic matter: artefacts, definitions, and functions. Geoderma 113, 187–209. doi:10.1016/S0016-7061(02)00361-0

Zubarev, R.A., Makarov, A., 2013. Orbitrap Mass Spectrometry. Anal. Chem. 85, 5288–5296. doi:10.1021/ac4001223

Zwetsloot, M.J., Kessler, A., Bauerle, T.L., 2018. Phenolic root exudate and tissue compounds vary widely among temperate forest tree species and have contrasting effects on soil microbial respiration. New Phytol. 218, 530–541. doi:10.1111/nph.15041

A   Appendix

In the following, supplemental tables, figures and explanations („notes") are presented (in this order). They refer to the main text.

Tables

Table A-1 Instrument settings of the two FTMS methods. Orbitrap-specific setting of Automatic Gain ControlTM and S-Lens RF level were adjusted at 1xE6 and 70%.

Conditions	FT-ICR MS	Orbitrap
Ionization mode	ESI Negative	ESI Negative
Flow [µl*min^{-1}]	4	7
Accumulation time [ms]	100	Max. 100 [1]
DOC [ppm]	10	20
Scan range [m/z]	115 – 2000	115 – 2000
Scans [n]	500	300
Source/ Capillary Temp. [°C]	200	275
Source fragmentation [eV]	40	40
Spray voltage [kV]	4.5	2.65
Transient length [s]	2.1	0.8

[1] Due to Automatic Gain Control (AGC).

Table A-2 Effect of the variation of a single factor (columns) on the different responses (with increasing factor level). Arrows show either a positive or a negative relation, no effect is marked by "o". Colors correspond to the ideal conditions (e.g., high percentage of assigned peaks, small variation coefficient): Green marks positive, yellow negative and blue neutral evaluation. Source fragmentation, inject time and sheath gas were the most influential factors but not checked for interactions. Variation coefficient of triples was always very low (i.e., good) but increased substantially by use of source fragmentation and high inject times.

	DOC	Flow rate	Source fragm.	S-Lens	Spray voltage	Sheath gas	Capillary temp.	Inject time
Unit	ppm	µl*min^{-1}	eV	%	kV	a.u.	°C	ms
Levels	10, 20, 30	5, 7, 10	0, 25, 50	30, 50, 70	2.5, 3.5, 4.5	0, 10, 20	225, 275, 325	100, 200, 500, 700
Number of peaks	↑	↑	o	o	↓	o	↑	o
% assigned formulae	o	o	↑↑	o	o	↑	o	↓↓
Similarity of data with ICR method (Pearson's r)	o	↓	↑↑	↑	o	↑↑	o	↓↓
% triple detected formulae in triplicates	o	o	↓	o	↑	↑	↓↓	↓↓
Variation coefficient of triples	o	↓	↑↑	↑	↑	o	o	↑↑

Table A-3 Molecular indices calculated from FTMS data.

Index	Explanation	Calculation/ definition	Reference of use
H/C	Atomic ratio of hydrogen to carbon in a molecular formula	H/C	Kew et al., 2017; Kim et al., 2003
O/C	Atomic ratio of oxygen to carbon in a molecular formula	O/C	
DBE	Double Bond Equivalents	$DBE=1+0.5*(2*C-H+N+P)$	Koch and Dittmar, 2016, 2006
AImod	Aromaticity index	$AImod=[1+C-0.5*O-S-0.5*(N+P+H)]/C-0.5*O-N-S-P$	
DBE/C	Carbon-normalized DBE	DBE/C	Lavonen et al., 2015; Roth et al., 2013
DBE-O	Oxygen-corrected DBE (sometimes half oxygen number)	DBE-O; sometimes also DBE-0.5*O	Herzsprung et al., 2014; Raeke et al., 2017; Roth et al., 2013
NOSC	Nominal Oxidation State of Carbon	$NOSC=4-[(4*C+1*H-3*N-2*O-2*S)/C]$	Riedel et al., 2013; other versions proposed by Boye et al., 2017; Kroll et al., 2011
BC	Polycyclic, condensed aromates, such as "Black Carbon"	$AImod \geq 0.66$	Modified from Šantl-Temkiv et al., 2013; other examples are given in e.g. D'Andrilli et al., 2015; Kellerman et al., 2014; Rossel et al., 2016; Seidel et al., 2014; Simon et al., 2019
PP	Polyphenols	$0.5 \geq AImod < 0.66$	
Prefix to BC& PP	„lw" – very low molecular weight „hw" – higher molecular weight	Additional constraint: C < 15 Additional constraint: $C \geq 15$	
HU	Highly unsaturated compound	$AImod \geq 0.5$; H/C <1.5; O/C < 0.9	
UA	Unsaturated aliphatics	$1.5 \geq H/C < 2$; O/C < 0.9; N = 0	
Prefix to PP, HU & UA	„or" – rich in oxygen „op" – poor in oxygen	Additional constraint: O/C >0.5 Additional constraint: $O/C \leq 0.5$	
SFA	Saturated, O-containing compound, such as fatty acids	$H/C \geq 2$; O/C < 0.9	
SUG	Very high O content, such as sugars	$O/C \geq 0.9$	
PEP	Unsaturated, O- and N-containing compound, such as peptides	$1.5 \geq H/C < 2$; O/C < 0.9; N > 0	
CHO, etc.	Formulae classified according to heteroatom content	Count formulae with only O, one N, one S, etc.	E.g., Pomerantz et al., 2011; Zhurov et al., 2013

Indices not included in this overview (and study) are nevertheless mentioned as alternatives here (amongst others): DOM lability index (D'Andrilli et al., 2015); index of terrestrial DOM/ marine DOM (Medeiros et al., 2016); DOM degradation indices (Flerus et al., 2012; Lechtenfeld et al., 2014; Mann et al., 2015); DOM diversity indices (Mentges et al., 2017); Kendrick mass defect of formulae (Hughey et al., 2001; Qi et al., 2016).

Table A-4 Information on standard compounds and solutions (structural formulae given in Figure 2-3).

ID	Standard compound	MW [Da]	Formula	Supplier	Weighed portion [mg]	Final concentration [ppm]
1	Vanillic acid	168.14	$C_8H_8O_4$	Sigma-Aldrich	1.98	200
2	4-Hydroxycinammic acid	164.04	$C_9H_8O_3$	Sigma-Aldrich	3.91	200
3	Gallic acid	170.12	$C_7H_6O_5$	Sigma-Aldrich	3.89	200
4	2-Methoxy-4-methylphenol	138.16	$C_8H_{10}O_2$	Sigma-Aldrich	10.9	200
5	3-Methoxyphenol	124.14	$C_7H_8O_2$	Sigma-Aldrich	13.1	200
64	2,3-Dimethoxy-5-methyl-1,4-benzoquinone	182.18	$C_9H_{10}O_4$	Alfa Aesar	2.5	200
7	Chlorogenic acid	354.31	$C_{16}H_{18}O_9$	Sigma-Aldrich	3.57	200
8	Ellagic acid	302.19	$C_{14}H_6O_8$	Sigma-Aldrich	0.99	< 124
9	6-o,p-Coumaryl-1,2-digalloylglucose	630.51	$C_{29}H_{26}O_{16}$	Sigma-Aldrich	0.35	39
10	Catechin	290.27	$C_{15}H_{14}O_6$	Sigma-Aldrich	1.35	100
11	Epigallocatechin gallate	458.37	$C_{22}H_{18}O_{11}$	Santa Cruz	0.98	100
12	Spiraeoside	464.38	$C_{21}H_{20}O_{12}$	Carl Roth	0.85	100
13	Isoquercetin	464.38	$C_{21}H_{20}O_{12}$	Santa Cruz	0.49	55
14	Myricitrin	464.38	$C_{21}H_{20}O_{12}$	Sigma-Aldrich	0.31	33

Table A-5 Instrument settings for fragmentation experiments.

Method stage	Factor	Standard compounds	DOM samples
Sample	DOC [ppm] Solvent Flow [µl*min^{-1}]	Max. 200, see Table S-1 50/50 MeOH/ H_2O 10	100 50/50 MeOH/ H_2O 7
Electrospray ionization	Ionization mode Source fragmentation [eV] Needle position Sheath gas [a.u.] Aux gas [a.u.] Sweep gas [a.u.] Spray voltage [kV] Capillary Temp. [°C]	Negative 0 Variable Variable Variable 0 Variable 275	Negative 40 D 25 0 0 2.65 275
Ion optics	S-Lens RF level [%] Multipole 00 offset [V] Lens 0 [V] Multipole 0 offset [V] Lens 1 [V] Multipole 1 offset [V] Multipole RF Amplitude [Vp-p] Front Lens [V]	Variable 1.1 3.2 9.4 17.3 13.8 800 10.3	70 1.0 3.2 9.4 17.2 13.2 792 10.0
Tandem MS	Act Q Act time [ms] Isolation window [amu] Modes and energy levels	0.25 0.1 1 CID: 15, 20, 25	0.25 0.1 1 CID: 15, 20, 25
MS Detection	Max. Inject time [ms] Automatic Gain Control^{TM} Scans per MS^2 experiment [n] Resolution Transient length [s] Profile mode Scan range [m/z]	5 5E4 50 240.000 0.8 Reduced Variable	2 5E4 150 240.000 0.8 Reduced Variable

Table A-6 Recalibration peaks used for standard compound Orbitrap tandem MS measurements. Compound #1 – #6 were only recalibrated by precursor ion exact m/z. References: [1] Ncube et al., 2014; [2] Mullen et al., 2003; [3] Fischer et al., 2011; [4] Engström et al., 2015; [5] Wyrepkowski et al., 2014; [6] Rockenbach et al., 2012; [7] Gu et al., 2003; [8] Miketova et al., 2000; [9] Yuzuak et al., 2018; [10] Fabre et al., 2001; [11] Saldanha et al., 2013.

ID	Standard compound	Precursor exact *m/z*	Product ions used as recal peaks, exact m/z (Formula)	Reference
7	Chlorogenic acid	353.088	191.0561 (C7H11O6), 179.035 (C9H7O4), 109.0295 (C6H5O2)	[1]
8	Ellagic acid	300.999	229.0143 (C12H5O5), 185.0244 (C11H5O3), 145.0296 (C9H5O2)	[2, 3, 4, 5]
9	6-o,p-Coumaryl-1,2-digalloylglucose	629.115	459.0933 (C22H19O11), 465.0675 (C20H17O13), 169.0142 (C7H5O5), 163.0401 (C9H7O3)	[6, 7]
10	Catechin	289.072	109.0295 (C6H5O2)	[6, 7, 8, 9]
11	Epigallocatechin gallate	457.078	169.0142 (C7H5O5)	[7, 8]
12	Spiraeoside	463.088	301.0354 (C15H9O7), 178.9986 (C8H3O5), 107.0139 (C6H3O2)	[10]
13	Isoquercetin	463.088	301.0354 (C15H9O7), 178.9986 (C8H3O5), 151.0037 (C7H3O4), 107.0139 (C6H3O2)	[4, 10]
14	Myricitrin	463.088	316.0225 (C15H8O8), 317.0303 (C15H9O8), 178.9986 (C8H3O5), 151.0037 (C7H3O4)	[10, 11]

Table A-7 Precursor and major product ions of the 14 standard compounds. The deprotonated precursor ion form was always dominant, except for compound 6 where the radical anion form dominated. Numbers in brackets indicate %-ion abundance relative to base peak (=100%) and the respective normalized CID energy at which mass spectra were acquired. In some cases further MS3 experiments were conducted at CID25.

ID	Standard compound	Formula	Precursor ion (m/z)	Product ions (m/z)
1	Vanillic acid	C8H8O4	167.0350 (35; at CID 25)	**152.0115** (92); **123.0452** (100); 109.0925 (<1); **108.0217** (18); 95.0503 (<1)
2	4-Hydroxycinammic acid	C9H8O3	163.0401 (25; at CID 25)	145.0296 (<1); 121.0296 (<1); **119.0295** (100); 93.0346 (<1)
3	Gallic acid	C7H6O5	169.0142 (16; at CID 25)	**125.0244** (100)
4	Creosol	C8H10O2	137.0608 (8; at CID 25)	**122.0374** (100); **109.0295** (2); 95.0503 (<1); 95.0139 (<1); 93.0346 (<1)
5	m-Guaiacol	C7H8O2	123.0452 (8; at CID 25)	**108.0217** (100); **95.0139** (1)
6	2,3-Dimethoxy-5-methyl-1,4-benzoquinone	C9H10O4	182.0585 (6; at CID20)	**167.0350** (100); 152.0115 (<1) MS³ (167, at CID20): 167.035 (1); **152.0115** (100); **139.0401** (3); **125.0245** (1); 121.0296 (<1)
7	Chlorogenic acid	C16H18O9	353.0878 (7; at CID20)	**191.0561** (100); **179.0350** (4); 161.0245 (<1); 109.0295 (<1); 99.0451 (<1)
8	Ellagic acid	C14H6O8	300.9990 (100; at CID 25)	**257.0092** (5); **229.0143** (5); **201.0193** (1); **185.0244** (3); 163.0401 (<1); 161.0245 (<1); 145.0296 (<1)
9	6-o,p-Coumaryl-digalloyl-Glucose	C29H26O16	629.1148 (2; at CID 25)	**477.1039** (8); **465.0675** (100); **459.0933** (48); **313.0565** (3); **271.0459** (5); 193.0142 (<1); 187.0401 (<1)
10	Catechin	C15H14O6	289.0718 (22; at CID 25)	**271.0612** (3); **247.0612** (5); **245.0820** (100); **231.0299** (6); **227.0714** (2); **205.0506** (35); **203.0714** (8); 188.0479 (1); 187.0401 (1); **179.035** (15); **167.035** (2); **165.0194** (4); 163.0401 (<1); 162.0323 (<1); **161.0609** (2); 161.0245 (<1); **151.0401** (2); **125.0244** (5); 123.0452 (<1); 121.0296 (<1); **109.0295** (2); 99.0451 (<1); 93.0346 (<1)
11	Epigallocatechin Gallate	C22H18O11	457.0776 (8; at CID 20)	**413.0879** (2); **331.0459** (95); **319.0458** (5); **305.0666** (33); **287.0561** (10); **275.0561** (3); **269.0455** (5); **193.0142** (12); **169.0142** (100)
12	Spiraeoside	C21H20O12	463.0882 (3; at CID 20)	**301.0354** (100) MS³ (301, at CID 25): 301.0354 (35); 300.0275 (<1); **273.0405** (10); **257.0455** (9); **229.0506** (2); **193.0142** (4); **178.9986** (100); **151.0037** (82); **121.0296** (1); **107.0138** (3)
13	Isoquercetin	C21H20O12	463.0882 (1; at CID 25)	**343.0459** (2); **301.0354** (100); **300.0275** (22) MS³ (301, at CID25): 301.0354 (32); 300.0275 (<1); **283.0248** (3); **273.0405** (11); **257.0455** (9); **255.0299** (1); **239.0350** (2); **229.0506** (3); **211.0401** (1); **193.0142** (4); **178.9986** (100); **151.0037** (88); **121.0296** (2); **107.0138** (4)
14	Myricitrin	C21H20O12	463.0882 (2; at CID 25)	**359.0408** (2); **337.0564** (1); **317.0303** (50); **316.0225** (100); **178.9986** (3)

Table A-8 Results of standard compound's tandem MS data analysis with Sirius (Böcker and Dührkop, 2016; Rasche et al., 2012; for product ion annotation and fragmentation tree generation) and CSI:Finger ID (Dührkop et al., 2015; for structure prediction by comparison of fragmentation trees).

ID	Standard compound/ neutral molecular formula	CID Levels	Precursor	Sirius: Peaks and assigned formulae	Sirius: Fragmentation tree	CSI Finger ID result
1	Vanillic acid C8H8O4	10,15, 20,25	[M-H]- 167.03498	6 peaks, 83% peaks with assigned formula, 99.87 total explained intensity, -0.01 ppm absolute error (Median)	Correct formula = tree#1, Tree score 11.97 (100%), correct tree has lowest ppm error	Score 86.31%, 1st hit
2	4-Hydroxy-cinammic acid C9H8O3	10,20, 25	[M-H]- 163.04007	2 peaks, 100% peaks with assigned formula, 100 total explained intensity, 0 ppm absolute error (Median)	Correct formula = tree#1, Tree score 12.71 (99.94%), correct tree has lowest ppm error	no prediction possible
3	Gallic acid C7H6O5	10,15, 20,25	[M-H]- 169.01425	2 peaks, 100% peaks with assigned formula, 100 total explained intensity, -0.01 ppm absolute error (Median)	Correct formula = tree#1, Tree score 2.19 (98.63%), correct tree has lowest ppm error	no prediction possible
4	Creosol C8H10O2	10,20, 25	[M-H]- 137.06080	4 peaks, 100% peaks with assigned formula, 100 total explained intensity, 0.16 ppm absolute error (Median)	Correct formula = tree#1, Tree score 10.95 (99.95%), correct tree has lowest ppm error	Score 64.79%, 1st hit
5	m-Guaiacol C7H8O2	10,20, 25	[M-H]- 123.04515	3 peaks, 100% peaks with assigned formula, 100 total explained intensity, 0.23 ppm absolute error (Median)	Correct formula = tree#1, Tree score 7.4 (99.91%), correct tree has lowest ppm error	Score 58.04%, 2nd hit
6	2,3-Dimethoxy-5-methyl-1,4-benzoquinone C9H10O4	10,15, 20	[M]- 182.05846	3 peaks, 100% peaks with assigned formula, 100 total explained intensity, -0.01 ppm absolute error (Median)	Correct formula = tree#2, Tree score 5.95 (41.87%), correct tree has lowest ppm error	Score 57.32% (wrong isomer)
7	Chlorogenic acid C16H18O9	10,15, 20	[M-H]- 353.08781	6 peaks, 100% peaks with assigned formula, 100 total explained intensity, -0.34 ppm absolute error (Median)	Correct formula = tree#1, Tree score 7.15 (99.28%), correct tree has lowest ppm error	Score 89.60%, 1st hit
8	Ellagic acid C14H6O8	10,20, 25,30, 35,40	[M-H]- 300.99899	55 peaks, 85% peaks with assigned formula, 99.25 total explained intensity, -0.1 ppm absolute error (Median)	Correct formula = tree#2, Tree score 54.17 (7.71%), correct tree has lowest ppm error	Score 80.83%, 1st hit
9	6-op-Coumaryl-digalloyl-Glucose C29H26O16	10,15, 20,25	[M-H]- 629.11481	15 peaks, 87% peaks with assigned formula, 99.6 total explained intensity, -0.19 ppm absolute error (Median)	Correct formula = tree#1, Tree score 19.53 (26.29%), correct tree has lowest ppm error	Score 73.33 %, 1st hit
10	Catechin C15H14O6	10,15, 20,25, 30	[M-H]- 289.07176	41 peaks, 98% peaks with assigned formula, 99.94 total explained intensity, -0.03 ppm absolute error (Median)	Correct formula = tree#1, Tree score 59.49 (100%), correct tree has lowest ppm error	Score 82.12% (wrong isomer)
11	Epigallocatechin Gallate C22H18O11	10,15, 20	[M-H]- 457.07764	18 peaks, 67% peaks with assigned formula, 98.34 total explained intensity, 0.25 ppm absolute error (Median)	Correct formula = tree#1, Tree score 27.55 (68.78%), correct tree close to lowest ppm error	Score 84.36 %, 1st hit
12	Spiraeoside C21H20O12	10,15, 20	[M-H]- 463.08820	5 peaks, 40% peaks with assigned formula, 98.86 total explained intensity, -0.01 ppm absolute error (Median)	Correct formula = tree#1, Tree score 4.87 (35.26%), correct tree has lowest ppm error	no prediction possible
13	Isoquercetin C21H20O12	10,15, 20,25	[M-H]- 463.08820	9 peaks, 78% peaks with assigned formula, 99.61 total explained intensity, 0.24 ppm absolute error (Median)	Correct formula = tree#1, Tree score 4.65 (56.76%), correct tree has lowest ppm error	Score 92.25 %, 1st hit
14	Myricitrin C21H20O12	10,15, 20,25	[M-H]- 463.08820	12 peaks, 83% peaks with assigned formula, 99.5 total explained intensity, 0.26 ppm absolute error (Median)	Correct formula = tree#1, Tree score 12.79 (95.61%), correct tree has lowest ppm error	Score 86.90 %, 1st hit

Table A-9 List of all 50+5 MDs extracted from the standard compound dataset covering several types of aromatic structures (Figure 2-3). Non-indicative MDs also reported for DOM (Table S-1) are marked in the explanation column with [DOM]. Five MDs were added without detection in the tandem MS data of the standard compounds to enable their search in the DOM data (thus the final number of 55). They are indicated by [ADD] and included the neutral loss analogs of precursor ions of compounds #1, #4, #8 and #10, and the common product ion of compounds #12 and #13 (originating from a sugar loss: neutral molecular formula C6H10O5) used for MS3 experiments. Contribution of MS3 data is marked with an asterisk (*) at the compound ID. Compound identifiers are put in brackets if the MD was detected below 1% relative intensity (based on base peak) across three CID levels. MD's that contributed only below <1% were only taken into account if detected for more than one compound. Occurrence refers to matches across 159 precursor peaks investigated. Tables 1 and 2 in the main text include only those MDs that were found in DOM with the MD matching approach. Eq., equivalent; Comb., combination.

Formula	Exact MD	Compound ID	Explanation	Occur.
$CH_3^\bullet$	15.02347	1, 4, 5, 6, 6*	Methyl radical, loss from radical ion on (6)	19
H_2O	18.01056	(2), 10, 13*, (14)	Water [DOM]	64
CO	27.99491	(4), 6*, (8), 12*, 13*	Formyl transf./ Carbon Monooxide [DOM]	44
C_2H_4	28.03130	4, 5	β-oxidation/ fatty acid synthesis [DOM]	44
C_2H_2O	42.01056	(2), (4), 6*, 10	Hydroxypyruvic acid/ -H_2O [DOM]	13
CO_2	43.98983	1, 2, 3, (7), 8, 10, 11, 12*, 13*	Carbon dioxide/ Carboxyl group [DOM]	96
CH_2O_2	44.99765	(2), (8)	Formic acid equivalent, radical	2
CH_2O_2	46.00548	(6*), 13, (13*)	Formic acid equivalent	20
C_3H_6O	58.04186	10	Acetone eq.; comb. C2H2O (ethenone) + CH4	14
$C_2H_4O_2$	59.01330	1, (10)	Acetic acid, radical	10
CH_2O_3	62.00039	10, 13*	Comb., CO_2 + H_2O [DOM]	43
C_2O_3	71.98474	(1), 8, (10), 12*, 13*	Comb., CO_2 + CO [DOM], Carbon Suboxide	31
$C_4H_4O_2$	84.02113	10	Combination, C3O2 (carbon suboxide) + CH4	6
$C_3H_2O_3$	86.00039	(1), 10	Combination, C3O2 (carbon suboxide) + H2O	13
$C_2H_2O_4$	89.99531	(10), 13*	Oxalic acid equivalent	20
C_3O_4	99.97966	8	Comb., CO_2 + 2x CO	15
$C_4H_6O_3$	101.02387	10	Radical loss from ion, not matched	0
$C_4H_6O_3$	102.03169	10	Comb., C4H4O2 + H2O (Sirius)	13
$C_4H_8O_3$	104.04734	14	Hydroxybutyric acid equivalent	16
$C_6H_4O_2$	108.02113	12*, 13*	Benzoquinone equivalent	7
$C_6H_6O_2$	110.03678	10	Benzenediol eq.; comb., C3O2 + CH4 + C2H2 (Sir.)	2
$C_4H_2O_4$	113.99531	(8), (10)	Butynedioic acid equivalent	9
C_3O_5	115.97457	8	Comb., 2x CO_2 + CO [DOM]	18
$C_4H_8O_4$	120.04226	13	Tetrose equivalent	15
$C_7H_6O_2$	122.03678	10, 12*, 13*	Loss from flavonols; Comb. on (10): C3O2 + C4H6	2
$C_7H_8O_2$	124.05243	10, Precursor (5)	3-Methoxyphenol (m-Guaiacol) unit	12
$C_6H_6O_3$	126.03169	(10), 11, 14	Phloroglucinol unit	11
$C_5H_4O_4$	128.01096	10	Comb., C3H4O2 + C2O2 (Sirius)	8
$C_7H_6O_3$	138.03169	10, 11, (13*)	Comb., C6H6O2 + CO (Sirius)	7
$C_8H_{10}O_2$	138.06808	Precursor (4)	[ADD] Creosol unit	7
$C_6H_{10}O_4$	146.05791	14	Sugar unit	8
$C_6H_{12}O_4$	147.06573	14	Sugar unit, radical form	1
$C_8H_6O_3$	150.03169	12*, 13*	Specific loss from flavonols	16

Table A–9: continued

Formula	Exact MD	Compound ID	Explanation	Occur.
$C_5H_4O_4$	128.01096	10	Comb., C3H4O2 + C2O2 (Sirius)	8
$C_7H_6O_3$	138.03169	10, 11, (13*)	Comb., C6H6O2 + CO (Sirius)	7
$C_8H_{10}O_2$	138.06808	Precursor (4)	[ADD] Creosol unit	7
$C_6H_{10}O_4$	146.05791	14	Sugar unit	8
$C_6H_{12}O_4$	147.06573	14	Sugar unit, radical form	1
$C_8H_6O_3$	150.03169	12*, 13*	Specific loss from flavonols	16
$C_7H_4O_4$	152.01096	9, 11	Incomplete gallic acid unit; H2O retained	7
$C_9H_6O_3$	162.03169	7	Caffeoyl unit	9
$C_6H_{10}O_5$	162.05282	12, 13	Sugar unit	9
$C_6H_{12}O_5$	163.06065	(12), 13	Sugar unit, radical form	0
$C_9H_8O_3$	164.04734	9, 10, Precursor (2)	p-coumaric ac.; Comb. on (10): C7H6O3 + C2H2 (Sirius)	7
$C_8H_8O_4$	168.04226	Precursor (1)	[ADD] Vanillic acid unit	6
$C_7H_6O_5$	170.02152	9, 11, Precursor (3)	Gallic acid unit	5
$C_7H_{10}O_5$	174.05282	7, (14)	Quinic ac. (7)	7
$C_8H_4O_5$	180.00587	12*, 13*	Loss from flavonols	4
$C_9H_8O_4$	180.04226	(7), 10	Caffeic ac.; Comb. on (10): C7H8O2 + 2x CO (Sirius)	4
$C_8H_6O_5$	182.02152	11	Comb., C6H6O3 (e.g., Phloroglucinol) + C2O2 (Sirius)	4
$C_9H_{10}O_4$	182.05791	(7), (9)	Comb. on (9): Coumaryl + 2x H2O (Sirius)	5
$C_7H_8O_6$	188.03209	(9), 11	Comb., C7H6O5 (e.g., Gallic acid) + H2O (Sirius)	4
$C_9H_6O_5$	194.02152	12*, 13*	Loss from flavonols	6
$C_{13}H_{12}O_6$	264.06339	11	Degrad. Catechin C ring after loss A or B-ring	2
$C_{13}H_{16}O_7$	284.08960	(13), 14	Not matched	0
$C_{15}H_{12}O_6$	288.06339	11	Loss of Catechin, gallic ac. remains	0
$C_{15}H_{14}O_6$	290.07904	Precursor (10)	[ADD] Catechin unit	0
$C_{14}H_6O_8$	302.00627	Precursor (8)	[ADD] Ellagic acid unit	0
$C_{15}H_{10}O_7$	302.04265	Precursor (12*, 13*)	[ADD] Flavonol subunit	0
$C_{16}H_{12}O_7$	316.05830	9	Degrad. coumaryl subunit after gallic acid loss	0
$C_{18}H_{14}O_8$	358.06887	9	Degrad. sugar core after coumaryl/ galloyl loss	0

Table A-10 List of reported or proposed DOM MDs from MS1 studies (within-spectrum MDs or "mass spacings", as in refs [1], [2], [4] and [5]) and MS2 studies (tandem MS MDs, as presented in refs [6]–[9]). Occurrence refers to matches across 159 precursor peaks investigated. References: [1] Zhang et al., 2014; [2] Longnecker and Kujawinski, 2016; [3] Cortés-Francisco and Caixach, 2015; [4] Kunenkov et al., 2009; [5] Kujawinski and Behn, 2006; [6] Witt et al., 2009; [7] Osterholz et al., 2015; [8] Hawkes et al., 2018; [9] Pohlabeln and Dittmar, 2015.

Formula	Exact mass difference	Reference(s)	Explanation	Occur.
$C_{-1}H_{-2}O$	1.979265	[1, 2]	Acetic acid/ $-H_2O$ and $-CO_2$	2
H_2	2.01565	[1 - 5]	(De-)hydrogenation	2
C	12	[1 - 3]	Glyoxylic acid/ $-H_2O$ and $-CO_2$	0
OH_{-2}	13.979265	[2]	O/H_2 exchange	3
CH_2	14.01565	[1 - 5]	(De-)methylation	3
O	15.994915	[1 - 5]	(De-)hydroxylation/ Oxygen	5
CH_4	16.0313	[6]	Methane	4
H_2O	18.010565	[3, 6 - 8, a.o.]	Water	64
$CH_{-2}O$	25.979265	[1]	C=O insertion	1
CHN	27.010899	[1]	Formimino transfer	0
CO	27.994915	[1 - 4]	Formyl transfer/ Carbon Monooxide	44
C_2H_4	28.031300	[1 - 3]	β-oxidation/ fatty acid synthesis	44
$H_{-1}NO$	28.990164	[1]	Nitrosylation	4
CHO	29.00274	[3]	Formyl-group related	6
CH_2O	30.010565	[1, 3, 4]	Hydroxymethyl transfer	13
S	31.972072	[9]	Sulfur	15
CH_4O	32.026215	[7, 8]	Methanol	43
2x H_2O	36.021130	[7]	Combination	14
C_2H_2O	42.010565	[1, 4]	Hydroxypyruvic acid/ -H2O	13
C_3H_6	42.04695	[3]	Repeated (de-)methylation	13
CHNO	43.005814	[1]	Carbamoyl- or isocyanide transfer	7
CO_2	43.989830	[3, 6 - 8, a.o.]	Carbon dioxide/ Carboxyl group	96
C_2H_4O	44.026215	[2, 3]	Acetaldehyde analogon	85
C_3H_2O	54.010565	[4]	Propynal analogon	0
C_2O_2	55.98983	[1]	Glyoxylic acid/ -H2O	11
C_4H_8	56.0626	[3]	Repeated (de-)methylation	11
$CO_2 + H_2O$	62.000395	[6 - 8]	Combination	43
HNO_3	62.995617	[3]	Nitrate	0
SO_2	63.961902	[9]	Sulfur dioxide	1
C_4H_4O	68.026215	[2, 8]	Vinyl Ketene	3
$C_3H_2O_2$	70.005480	[4]	Propiolic acid analogon	6
$CO_2 + CO$	71.984745	[6]	Combination	31
$C_2H_3NO_2$	73.016379	[1]	Tryptophanase	1
$CO_2 + CH_4O$	76.016045	[7]	Combination	26
SO_3	79.956817	[9]	Sulfur trioxide	14
H_2SO_3	81.972467	[9]	Sulfurous acid	8
2x CO_2	87.979660	[3, 6 - 8]	Combination	44
2x $CO_2 + H_2O$	105.990225	[6 - 8]	Combination	26

Table A–10: continued

Formula	Exact mass difference	Reference(s)	Explanation	Occur.
$CO_2 + SO_2$	107.951732	[9]	Combination	6
2x $CO_2 + CO$	115.974575	[6]	Combination	18
2x $CO_2 + CH_4O$	120.005875	[7]	Combination	15
$CO_2 + SO_3$	123.946647	[9]	Combination	9
2x $CO_2 + 2 H_2O$	124.000790	[6]	Combination	12
3x CO_2	131.969490	[6, 7]	Combination	18
2x $CO_2 + H_2O + CO$	133.985140	[6]	Combination	13
3x $CO_2 + CH_4$	148.000790	[6]	Combination	10
3x $CO_2 + H_2O$	149.980055	[6, 7]	Combination	14
$C_7H_6O_4$	154.026610	[4]	Dihydroxyl-benzoic acid analogon	6
3x $CO_2 + CH_4O$	163.995705	[7]	Combination	7
3x $CO_2 + 2 H_2O$	167.990620	[6]	Combination	6
4x CO_2	175.959320	[6]	Combination	7
3x $CO_2 + H_2O + CO$	177.974970	[6]	Combination	5
4x $CO_2 + CH_4$	191.990620	[6]	Combination	6
4x $CO_2 + H_2O$	193.969885	[6]	Combination	4

Table A-11 Compilation of Orbitrap/ FT-ICR MS comparison studies for different natural organic matter (NOM) samples taken from [1] (Cortés-Francisco et al., 2011); [2] (Pomerantz et al., 2011); [3] (Remucal et al., 2012); [4] (Smith et al., 2012); [5] (Zhurov et al., 2013); [6] (Mangal et al., 2016); [7] (Hawkes et al., 2016); [8] This study; n.s., not stated; * improved analyzer & high transient length.

Ref.	Type of sample(s)	Instrument, ESI mode, m/z range	Evaluation parameters	Sample information
[1]	Standard mix of four ions and IHSS samples	LTQ Orbitrap/ Orbitrap Exactive, n.s., n.s.	Mass accuracy (drift) and precision, resolving power	None
[2]	Petroleum	LTQ Orbitrap XL, negative mode, 150 – 1050	Mass resolution, mass accuracy	Heteroatom class distribution, DBE distribution, cyclic/ acyclic ratio, PM degradation scale
[3]	IHSS samples, SRFA	Orbitrap Exactive, negative mode, (100) 290 – 600	Mass distribution, shared formulae	None
[4]	Fraction of a bio-oil produced by fast pyrolysis	LTQ Orbitrap Discovery, negative mode, 100 – 400	Mass distribution, mass resolution	Heteroatom class distribution, DBE distribution
[5]	Resin and maltene fraction of 2 crude oils	Orbitrap Elite*, positive mode, 200 – 1000	Resolving power, mass accuracy, spectral dynamic range	Kendrick mass defect analysis, heteroatom class subsets, heteroatom class distribution
[6]	IHSS samples, SRFA and PLFA	Orbitrap Q Exactive, negative mode, 200 – 1000	Mass distribution	Kendrick mass defect analysis, average DBE, average elemental composition
[7]	Mixture of a dystrophic lake sample and a marine sample	LTQ Velos Pro Orbitrap, negative mode, 150 – 2000	General performance, mass distribution, differentiation of composition changes	Average H/C and m/z values, critical mass differences, Bray-Curtis dissimilarities
[8]	Set of 17 SPE-DOM samples from a wide range of ecosystems	Orbitrap Elite*, negative mode, 200 – 650	Mass distribution, sample-specific loss of information, causes of information loss, reproducibility/ retrieval of biogeochemical trends	Ion abundance patterns, Van Krevelen patterns, critical mass differences, number of formulae and molecular group contribution, trends of molecular indices (DBE, NOSC, H/C, etc)

Table A-12 Properties of non-fragmented (CID 0) IPIMs (isolated precursor ion mixtures) at four masses (except number of fragments (CID25); determined at CID 25), and significance of correlation between nominal mass (m/z 241, 301, 361, 417) and CID energy (CID 15, 20, 25) with the respective metrics (p-value <0.05, significant), blue and red indicate positive/ negative correlation. Lighter colors or grey are chosen when significance levels are > 0.05. Brackets are put around obvious correlations: number of atoms in heavier molecules is higher, and precursor number sinks upon fragmentation.

Property	m/z 241	m/z 301	m/z 361	m/z 417	p-value m/z*	p-value CID
Precursors	33	37	43	44	0.026	(0.000)
Precursors ass.	21	30	34	40	0.043	(0.078)
Fragments$^{(CID25)}$	198	321	390	491	0.002	(0.000)
H/C$_{WA}$	0.91	0.94	0.98	0.99	0.032	0.003
O/C$_{WA}$	0.37	0.45	0.48	0.53	0.038	0.000
#C$_{WA}$	13.06	15.21	17.58	19.61	(0.020)	0.178
#H$_{WA}$	11.65	14.14	17.03	19.26	(0.000)	0.957
#O$_{WA}$	4.54	6.51	8.14	10.02	(0.003)	0.077
AI$_{mod,WA}$	0.53	0.47	0.42	0.39	0.010	0.004
DBE$_{WA}$	8.26	9.17	10.11	11.01	0.010	0.004
DBE-O$_{WA}$	3.72	2.66	1.97	0.99	0.003	0.000
NOSC$_{WA}$	-0.07	0.04	0.07	0.14	0.012	0.000

*In this case, equivalent to mass defect.

Table A-13 Overview of correlations (Pearson's r; red, negative correlation; blue, positive correlation) between key properties of the IPIM (representing the bandwidth of possible isomers behind a given exact precursor m/z) at m/z 241 (precursor ions with molecular formula = 20). Shown are descriptors of ionization and fragmentation behavior (i.e., initial intensity (Iabs, initial), fragmentation at different CID stages (Irel,loss) and number of matches to non-indicative MDs reported for DOM (Table S-1) and their relation to the precursor's m/z (here, equivalent to mass defect) and molecular formula (numbers of #C, #H and #O atoms, their atomic H/C and O/C ratios, the nominal oxidation state of carbons (NOSC) (Boye et al., 2017), number of oxygen-corrected double bond equivalents (DBE-O) (Herzsprung et al., 2014), and the number of CO_2 (0 – 4), H_2O (0 – 2), CO (0 – 1) and $C_7H_6O_4$ (0 – 1) (Kunenkov et al., 2009) losses inferred from non-indicative MDs and their combinations (Table S-1). Other molecular indices as double bond equivalent (DBE), aromaticity index (AImod) (Koch and Dittmar, 2016) and number of CH_2 losses (0 – 4) were tested but showed non-significant (ns) relationships in this analysis. Explanation of p-value notation: $p > 0.05$, "ns"; $0.05 \geq p > 0.01$, "*"; $0.01 \geq p > 0.001$, "**"; $p \leq 0.001$, "***".

	$I_{rel, loss, CID15}$	$I_{rel, loss, CID20}$	$I_{rel, loss, CID25}$	$I_{abs, initial}$	**Matches**
m/z	-0.59 **	-0.64 **	-0.68 **	-0.18 ns	-0.29 ns
# C	-0.63 **	-0.74 ***	-0.78 ***	-0.05 ns	-0.39 ns
# H	-0.5 *	-0.54 *	-0.59 **	-0.22 ns	-0.29 ns
# O	0.63 **	0.77 ***	0.77 ***	0.28 ns	0.62 **
H/C	-0.33 ns	-0.32 ns	-0.35 ns	-0.2 ns	-0.15 ns
O/C	0.68 **	0.78 ***	0.74 ***	0.16 ns	0.53 *
NOSC	0.61 **	0.66 **	0.69 ***	0.14 ns	0.35 ns
DBE-O	-0.29 ns	-0.4 ns	-0.36 ns	-0.01 ns	-0.31 ns
n CO₂	0.52 *	0.65 **	0.64 **	0.53 *	0.84 ***
n H₂O	0.33 ns	0.51 *	0.52 *	0.54 *	0.86 ***
n CO	-0.03 ns	0.12 ns	0.21 ns	0.69 ***	0.74 ***
n C₇H₆O₄					
$I_{rel, loss, CID15}$		0.94 ***	0.73 ***	0.06 ns	0.32 ns
$I_{rel, loss, CID20}$			0.88 ***	0.18 ns	0.5 *
$I_{rel, loss, CID25}$				0.25 ns	0.52 *
$I_{abs, initial}$					0.81 ***

Table A-14 Overview of correlations (Pearson's r; red, negative correlation; blue, positive correlation) between key properties of the IPIM (representing the bandwidth of possible isomers behind a given exact precursor m/z) at m/z 301 (precursor ions with molecular formula = 27). Shown are descriptors of ionization and fragmentation behavior (i.e., initial intensity (Iabs, initial), fragmentation at different CID stages (Irel,loss) and number of matches to non-indicative MDs reported for DOM (Table S-1) and their relation to the precursor's m/z (here, equivalent to mass defect) and molecular formula (numbers of #C, #H and #O atoms, their atomic H/C and O/C ratios, the nominal oxidation state of carbons (NOSC)(Boye et al., 2017), number of oxygen-corrected double bond equivalents (DBE-O)(Herzsprung et al., 2014), and the number of CO_2 $(0-4)$, H_2O $(0-2)$, CO $(0-1)$ and $C_7H_6O_4$ $(0-1)$(Kunenkov et al., 2009) losses inferred from non-indicative MDs and their combinations (Table S-1). Other molecular indices as double bond equivalent (DBE), aromaticity index (AImod)(Koch and Dittmar, 2016) and number of CH2 losses $(0-4)$ were tested but showed non-significant (ns) relationships in this analysis. Explanation of p-value notation: $p > 0.05$, "ns"; $0.05 \geq p > 0.01$, "*"; $0.01 \geq p > 0.001$, "**"; $p \leq 0.001$, "***".

	$I_{rel, loss, CID15}$	$I_{rel, loss, CID20}$	$I_{rel, loss, CID25}$	$I_{abs, initial}$	**Matches**
m/z	-0.47 *	-0.66 ***	-0.43 *	-0.12 ns	-0.22 ns
# C	-0.59 **	-0.85 ***	-0.87 ***	-0.06 ns	-0.35 ns
# H	-0.35 ns	-0.53 **	-0.3 ns	-0.12 ns	-0.17 ns
# O	0.64 ***	0.9 ***	0.83 ***	0.24 ns	0.54 **
H/C	-0.12 ns	-0.18 ns	0.1 ns	-0.08 ns	0.01 ns
O/C	0.64 ***	0.87 ***	0.77 ***	0.11 ns	0.45 *
NOSC	0.48 *	0.73 ***	0.56 **	0.03 ns	0.21 ns
DBE-O	-0.49 **	-0.64 ***	-0.78 ***	-0.1 ns	-0.41 *
n CO$_2$	0.59 **	0.62 ***	0.46 *	0.54 **	0.85 ***
n H$_2$O	0.36 ns	0.48 *	0.49 **	0.5 **	0.71 ***
n CO	-0.2 ns	-0.14 ns	-0.12 ns	0.44 *	0.21 ns
n C$_7$H$_6$O$_4$					
$I_{rel, loss, CID15}$		0.83 ***	0.55 **	0.15 ns	0.52 **
$I_{rel, loss, CID20}$			0.84 ***	0.25 ns	0.56 **
$I_{rel, loss, CID25}$				0.26 ns	0.47 *
$I_{abs, initial}$					0.81 ***

Table A-15 Overview of correlations (Pearson's r; red, negative correlation; blue, positive correlation) between key properties of the IPIM (representing the bandwidth of possible isomers behind a given exact precursor m/z) at m/z 361 (precursor ions with molecular formula = 30). Shown are descriptors of ionization and fragmentation behavior (i.e., initial intensity (Iabs, initial), fragmentation at different CID stages (Irel,loss) and number of matches to non-indicative MDs reported for DOM (Table S-1) and their relation to the precursor's m/z (here, equivalent to mass defect) and molecular formula (numbers of #C, #H and #O atoms, their atomic H/C and O/C ratios, the nominal oxidation state of carbons (NOSC)(Boye et al., 2017), number of oxygen-corrected double bond equivalents (DBE-O)(Herzsprung et al., 2014), and the number of CO_2 $(0-4)$, H_2O $(0-2)$, CO $(0-1)$ and $C_7H_6O_4$ $(0-1)$(Kunenkov et al., 2009) losses inferred from non-indicative MDs and their combinations (Table S-1). Other molecular indices as double bond equivalent (DBE), aromaticity index (AImod)(Koch and Dittmar, 2016) and number of CH_2 losses $(0-4)$ were tested but showed non-significant (ns) relationships in this analysis. Explanation of p-value notation: $p > 0.05$, "ns"; $0.05 \geq p > 0.01$, "*"; $0.01 \geq p > 0.001$, "**"; $p \leq 0.001$, "***".

	$I_{rel, loss, CID15}$	$I_{rel, loss, CID20}$	$I_{rel, loss, CID25}$	$I_{abs, initial}$	**Matches**
m/z	-0.58 ***	-0.6 ***	-0.3 ns	-0.1 ns	-0.24 ns
# C	-0.76 ***	-0.88 ***	-0.85 ***	-0.01 ns	-0.24 ns
# H	-0.49 **	-0.47 **	-0.13 ns	-0.11 ns	-0.21 ns
# O	0.84 ***	0.92 ***	0.81 ***	0.16 ns	0.4 *
H/C	-0.17 ns	-0.08 ns	0.25 ns	-0.12 ns	-0.1 ns
O/C	0.85 ***	0.9 ***	0.75 ***	0.03 ns	0.28 ns
NOSC	0.76 ***	0.74 ***	0.43 *	0.03 ns	0.2 ns
DBE-O	-0.51 **	-0.64 ***	-0.8 ***	-0.02 ns	-0.21 ns
n CO$_2$	0.45 *	0.51 **	0.43 *	0.71 ***	0.83 ***
n H$_2$O	0.26 ns	0.42 *	0.46 *	0.62 ***	0.79 ***
n CO	-0.01 ns	0.07 ns	0.09 ns	0.63 ***	0.6 ***
n C$_7$H$_6$O$_4$					
$I_{rel, loss, CID15}$		0.92 ***	0.66 ***	0.05 ns	0.28 ns
$I_{rel, loss, CID20}$			0.85 ***	0.21 ns	0.45 *
$I_{rel, loss, CID25}$				0.26 ns	0.44 *
$I_{abs, initial}$					0.92 ***

Table A-16 Overview of correlations (Pearson's r; red, negative correlation; blue, positive correlation) between key properties of the IPIM (representing the bandwidth of possible isomers behind a given exact precursor m/z) at m/z 417 (precursor ions with molecular formula = 34). Shown are descriptors of ionization and fragmentation behavior (i.e., initial intensity (Iabs, initial), fragmentation at different CID stages (Irel,loss) and number of matches to non-indicative MDs reported for DOM (Table S-1) and their relation to the precursor's m/z (here, equivalent to mass defect) and molecular formula (numbers of #C, #H and #O atoms, their atomic H/C and O/C ratios, the nominal oxidation state of carbons (NOSC)(Boye et al., 2017), number of oxygen-corrected double bond equivalents (DBE-O)(Herzsprung et al., 2014), and the number of CO_2 $(0-4)$, H_2O $(0-2)$, CO $(0-1)$ and $C_7H_6O_4$ $(0-1)$(Kunenkov et al., 2009) losses inferred from non-indicative MDs and their combinations (Table S-1). Other molecular indices as double bond equivalent (DBE), aromaticity index (AImod)(Koch and Dittmar, 2016) and number of CH2 losses $(0-4)$ were tested but showed non-significant (ns) relationships in this analysis. Explanation of p-value notation: $p > 0.05$, "ns"; $0.05 \geq p > 0.01$, "*"; $0.01 \geq p > 0.001$, "**"; $p \leq 0.001$, "***".

	$I_{rel, loss, CID15}$	$I_{rel, loss, CID20}$	$I_{rel, loss, CID25}$	$I_{abs, initial}$	**Matches**
m/z	-0.58 ***	-0.67 ***	-0.38 *	-0.2 ns	-0.36 *
# C	-0.64 ***	-0.79 ***	-0.78 ***	-0.19 ns	-0.43 *
# H	-0.49 **	-0.56 ***	-0.27 ns	-0.18 ns	-0.31 ns
# O	0.79 ***	0.88 ***	0.82 ***	0.37 *	0.63 ***
H/C	-0.23 ns	-0.24 ns	0.05 ns	-0.1 ns	-0.13 ns
O/C	0.8 ***	0.86 ***	0.75 ***	0.28 ns	0.55 ***
NOSC	0.67 ***	0.77 ***	0.53 **	0.19 ns	0.4 *
DBE-O	-0.42 *	-0.49 **	-0.65 ***	-0.18 ns	-0.35 *
n CO_2	0.72 ***	0.74 ***	0.61 ***	0.74 ***	0.89 ***
n H_2O	0.51 **	0.57 ***	0.56 ***	0.54 **	0.67 ***
n CO	0.13 ns	0.21 ns	0.18 ns	0.57 ***	0.53 **
n $C_7H_6O_4$	0.22 ns	0.26 ns	0.2 ns	0.84 ***	0.7 ***
$I_{rel, loss, CID15}$		0.89 ***	0.54 ***	0.44 *	0.69 ***
$I_{rel, loss, CID20}$			0.76 ***	0.46 **	0.69 ***
$I_{rel, loss, CID25}$				0.32 ns	0.52 **
$I_{abs, initial}$					0.92 ***

Figures
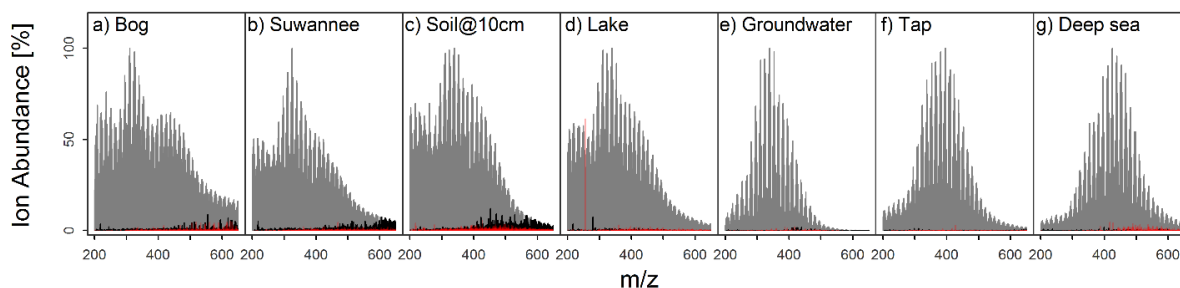
Figure A-1 Distribution of Orbitrap signals (formulae) for the samples of the subset used for detailed analyses (compare Fig.1 and Fig. 2 in main text). Shown are specific signals (red, specific s.s.), signals also encountered in the FT-ICR MS dataset but not in same samples (black, specific s.l.), and common signals (both datasets, and exact same sample (grey). It is obvious that common signals are by far the most prominent fraction for all sample types. In the higher mass range, unavoidable differences in tuning led to slight deviations in apparent sample composition (black signals with higher intensities, especially in a, b, c). The effect was most pronounced for samples with ion abundance maxima in the lower mass range and influenced multivariate separation of samples, too (e.g., leading to an offset between mean average m/z as derived from FT-ICR MS and Orbitrap data, Supplementary Figure 8; see also main text).
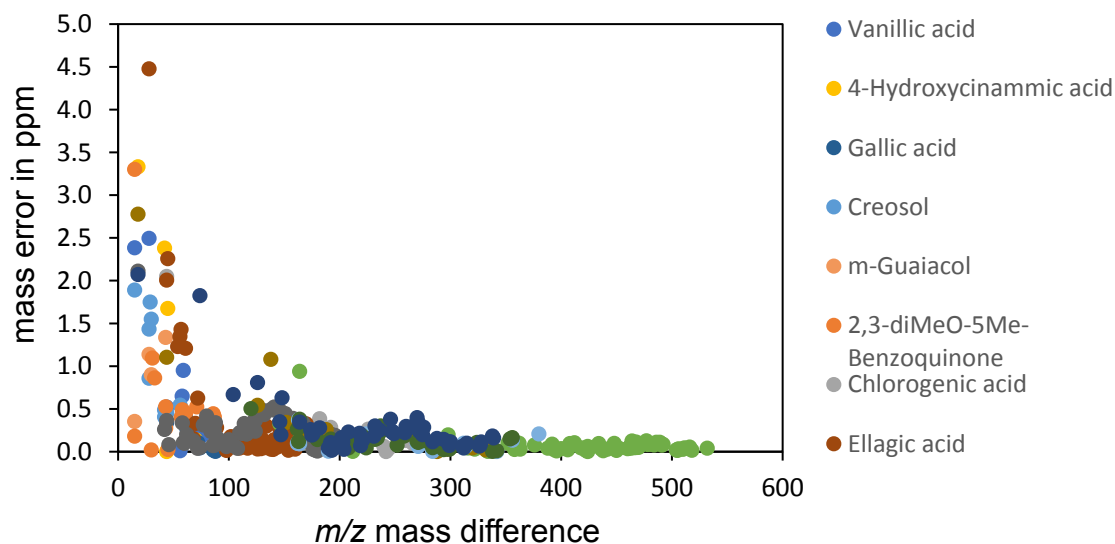
Figure A-2 Error assessment of standard compound MDs (deviation between measured MD and exact MD as predicted by molecular formula of precursor and product ions. Relative errors become large when the mass difference is small (Kujawinski and Behn, 2006).
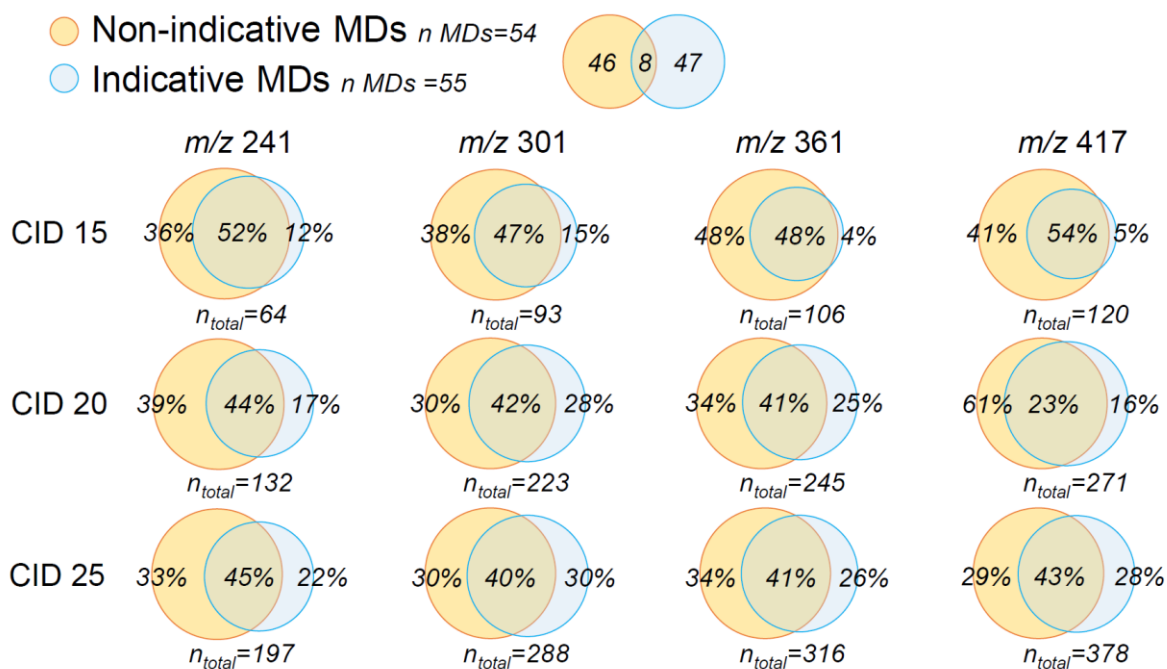
Figure A-3 Comparison of matches to the two MD lists (indicative and non-indicative) in dependence of nominal mass (m/z) and collision energy (CID), shown as Venn diagrams. ntotal designates the total number of MD matches at each CID stage for each IPIM (nominal mass, isobar). Percentages show the relative amount of unique or shared (overlap) matches between both lists. Note that Venn circles on top designate overlap in terms of absolute number of MDs between lists. Not all MDs were found in DOM.
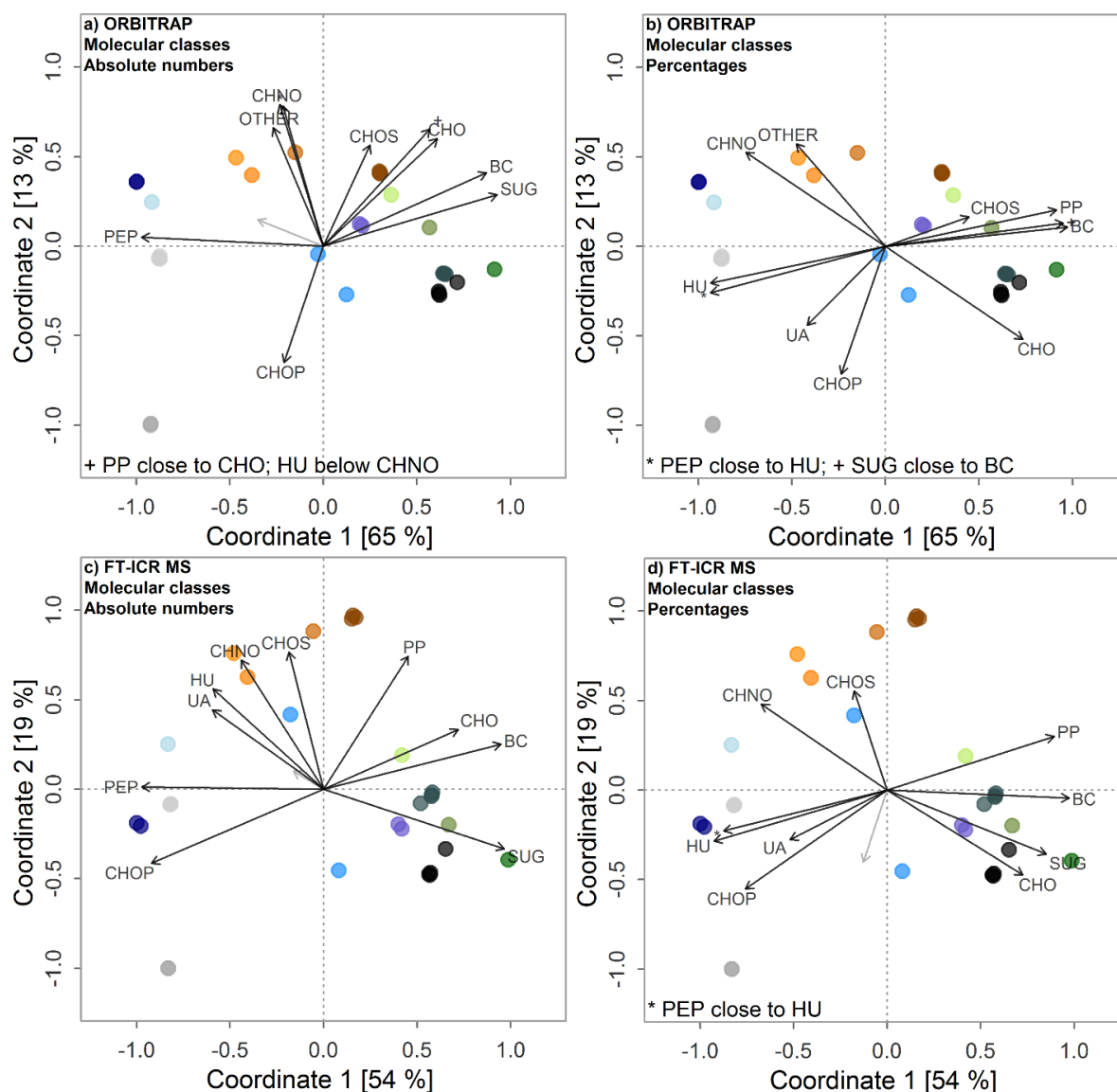
Figure A-4 Post-gradient fitting on ordination obtained by Orbitrap (upper panels) and FT-ICR MS (lower panels) for the molecular class data (Molecular groups/ Formula classes). Fitted gradients are based on percentage data (b, d) or absolute numbers (a, c). Molecular classes: BC, black carbon; PP, polyphenol; HU, highly unsaturated; UA, unsaturated aliphatics; SUG, carbohydrate; PEP, peptide. CHNO includes all Nitrogen-containing formulae (N1-4), and CHOS includes all Sulfur-containing formulae (S1-2). OTHER is the sum of CHOSP, CHNOS and CHNOP formulae.
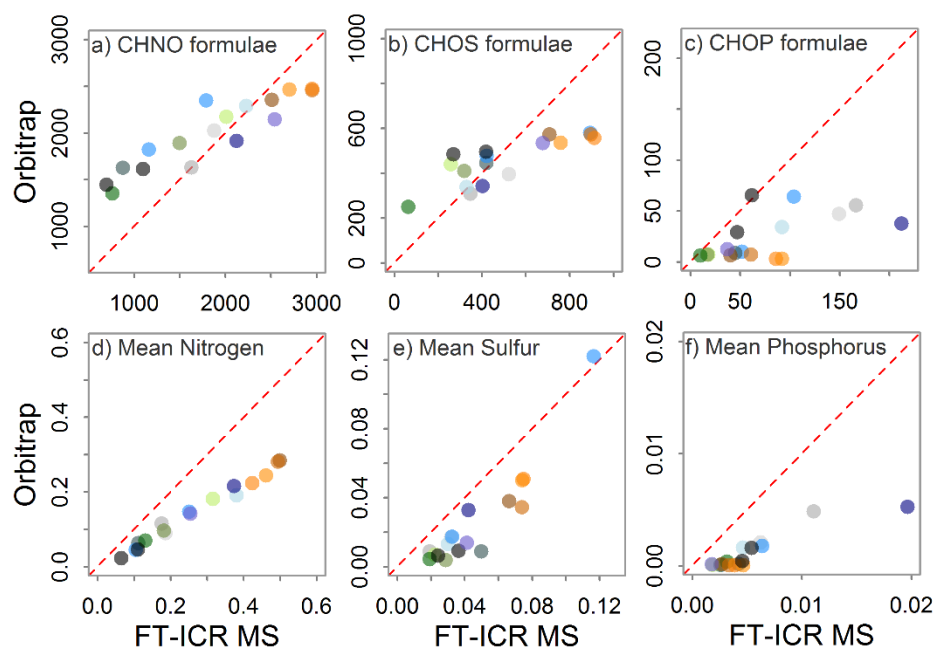
Figure A-5 Instrumental offsets due to detection of non-common signals by FT-ICR MS (due to better resolution), as shown by molecular class indices linked to N, S and P formulae, and based on the full dataset. The shown indices exerted a trend towards lower values in the Orbitrap due to limitations in resolving power (panels d, e and f) as described in the main text. The differences in tuning lead to a slight mass shift in the Orbitrap which overlays this general resolution effect (in panels a and b) and accounts for slightly higher numbers of simple CHNO and CHOS formulae detected by the Orbitrap in samples characterized by recent inputs of fresh and degrading organic matter.
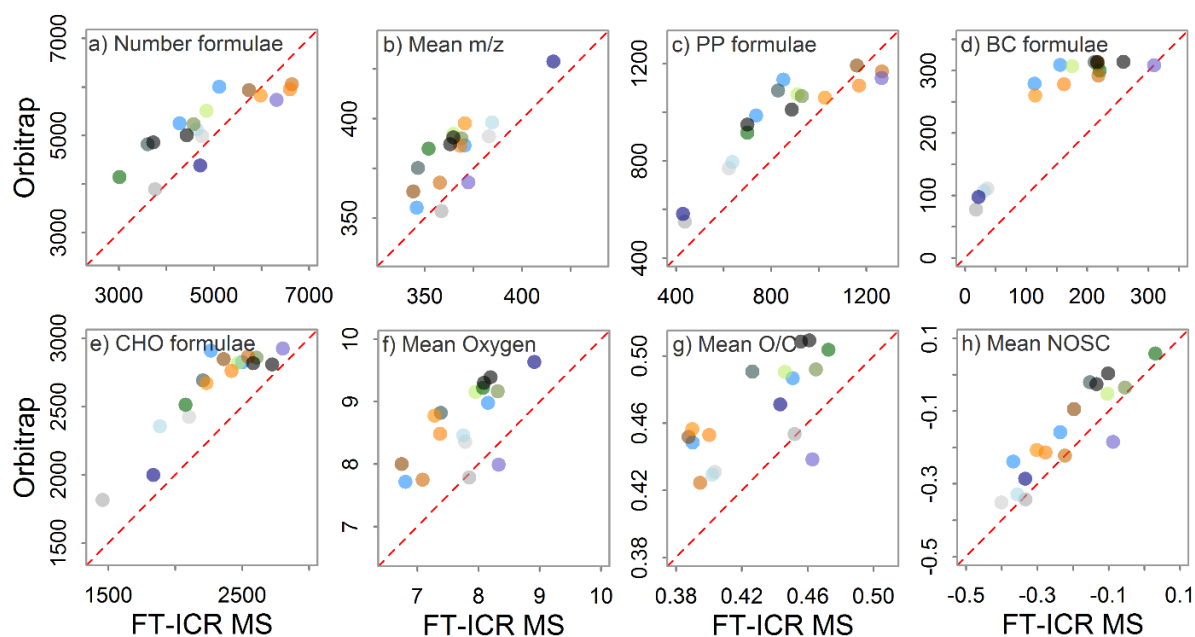
Figure A-6 Instrumental offsets due to detection of signals by Orbitrap in the higher mass range (due to differential architecture, tuning and method setup), shown by selected chemical and molecular class indices, and based on the full dataset. The shown indices exerted a trend towards higher values in the Orbitrap with a tendency to be more pronounced for samples connected to recent inputs of fresh and degrading organic matter. The observed higher number of compounds are related to unavoidable differences in instrumental response due to tuning (Orbitrap was tuned with IHSS sample, FT-ICR MS with NELHA). The Orbitrap-detected signals of mainly aromatic CHO and simple CHNO/ CHOS formulae in the higher mass range were also found by FT-ICR MS but not within the exact same samples; this difference being mainly an effect of tuning and not FT-ICR MS capabilities.
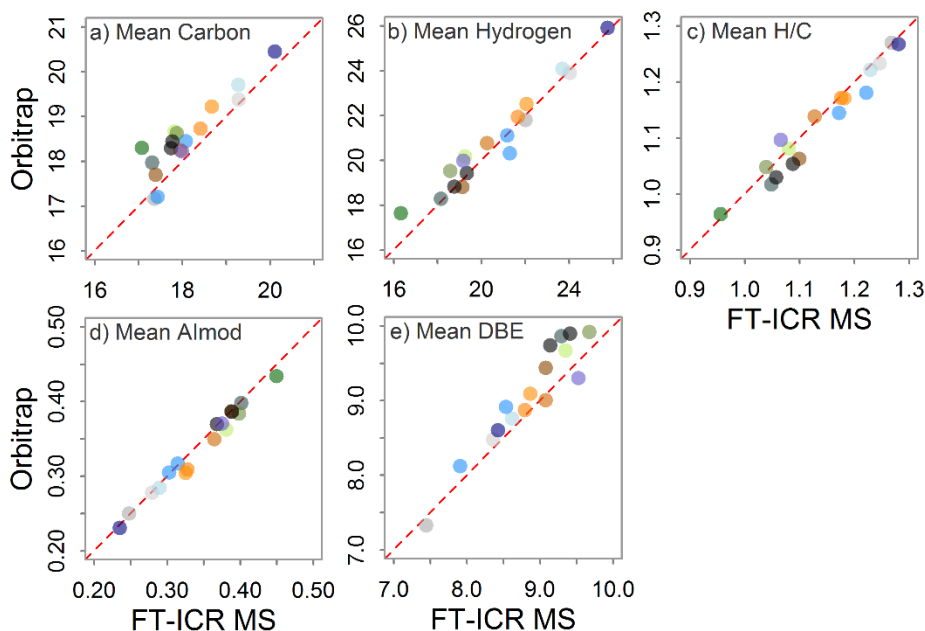
Figure A-7 Examples of chemical indices that showed no strong offsets between instruments, as based on the full dataset.
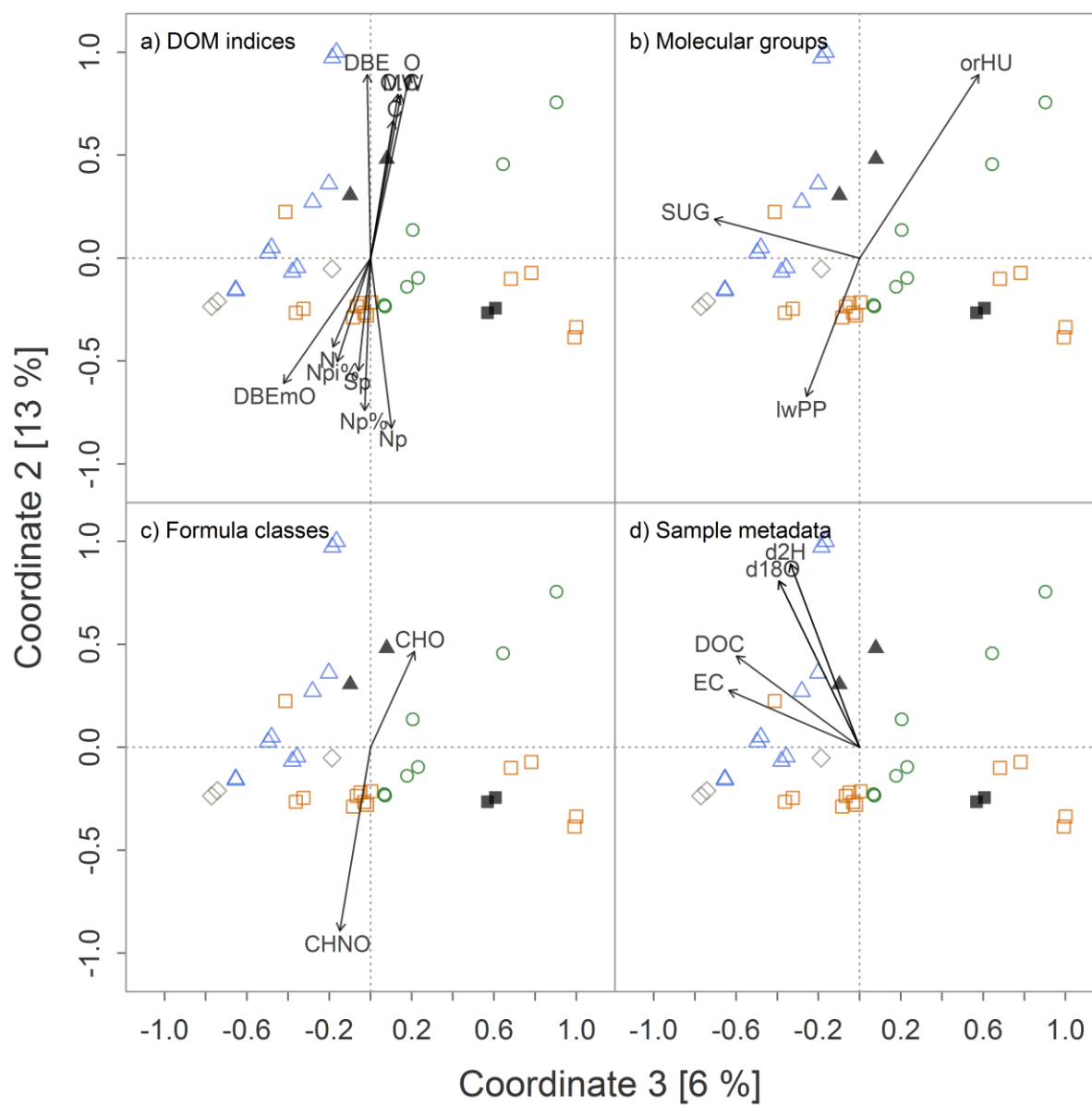
Figure A-8 Separation of the whole set of samples by coordinates 2 (13% explained variation) and 3 (6%) in a PCoA and subsequent envfit analysis.
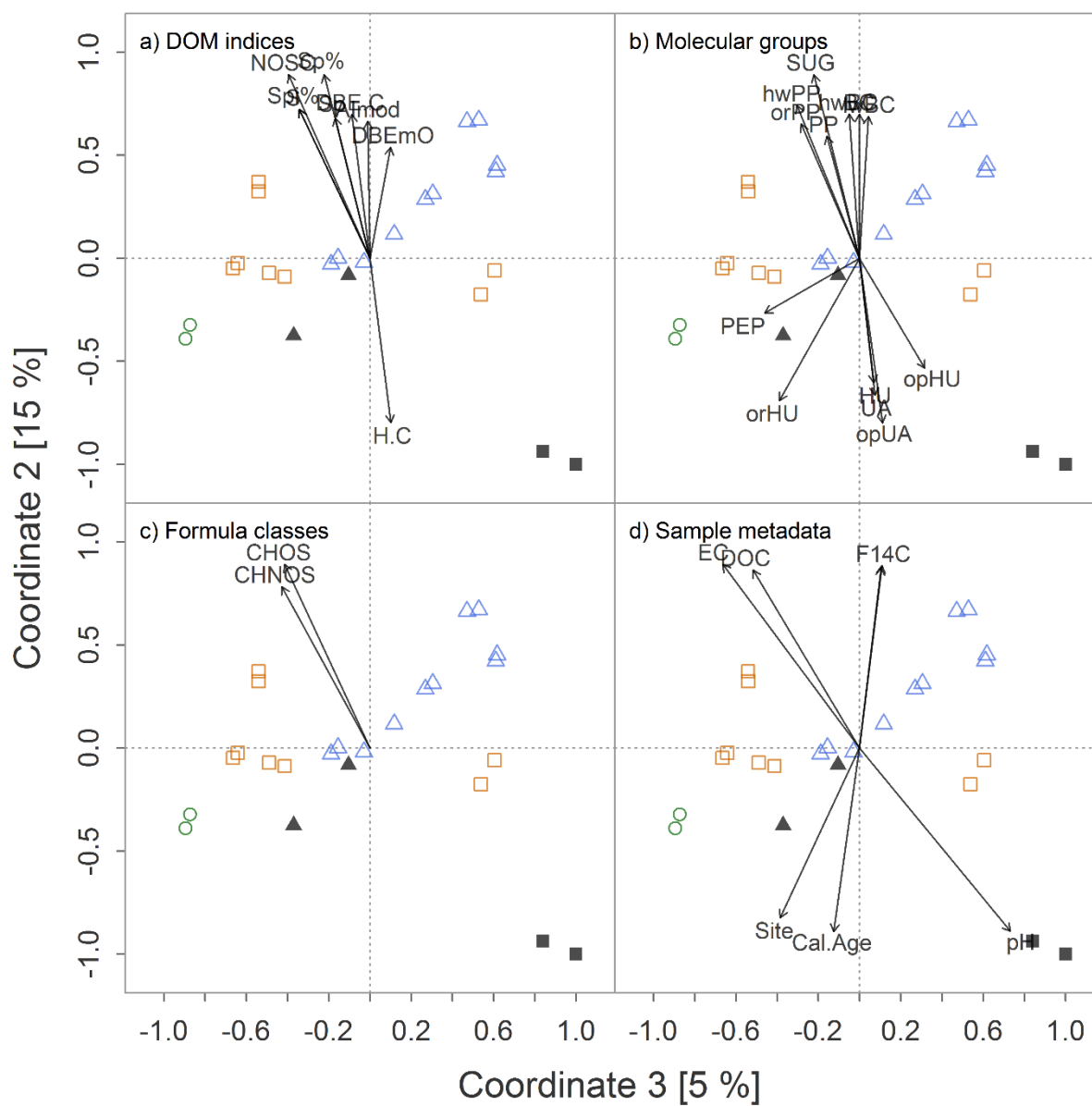
Figure A-9 Separation of the subset of samples measured successfully for ^{14}C content by coordinates 2 (15% explained variation) and 3 (5%) in a PCoA and subsequent envfit analysis.
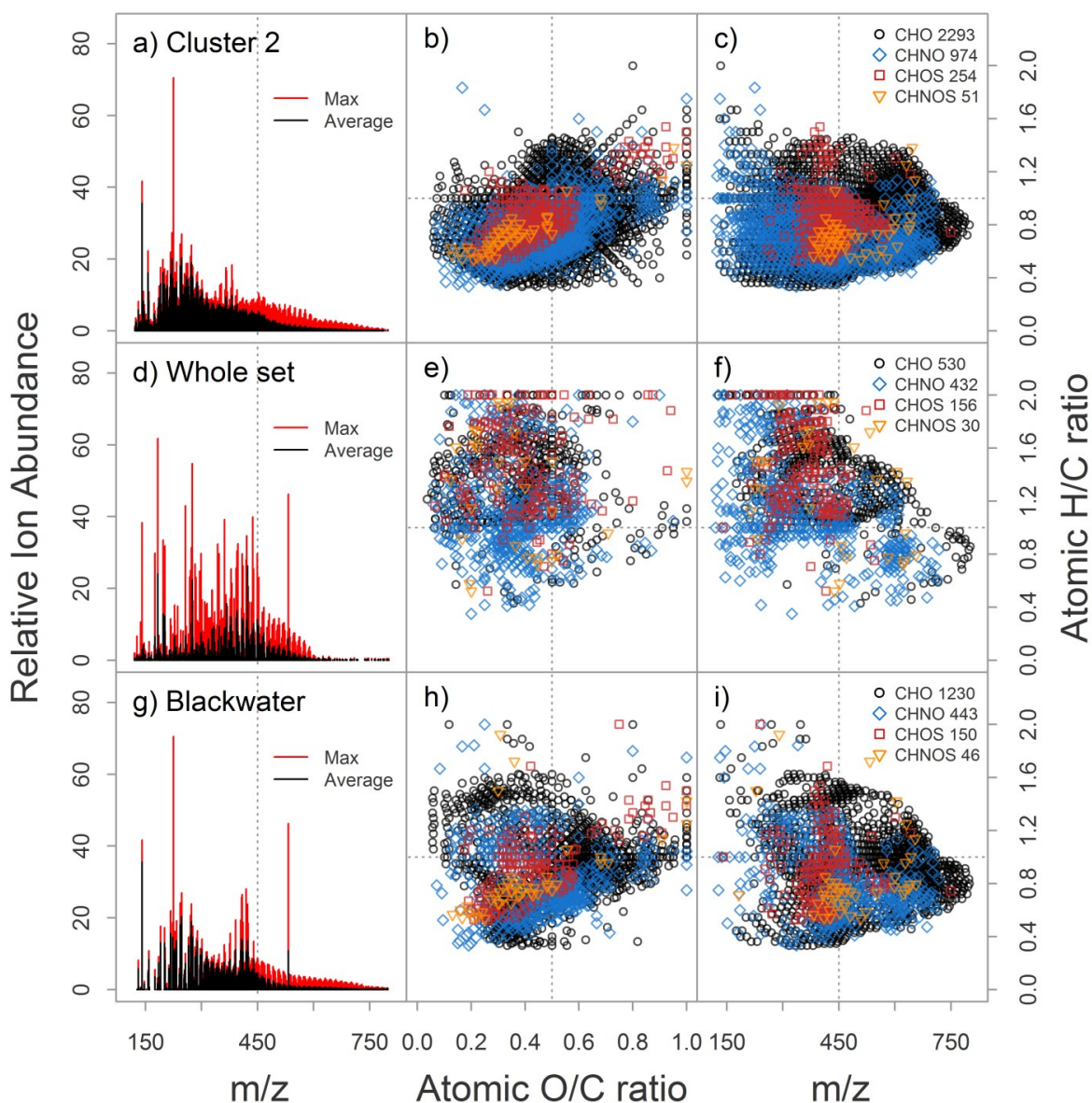
Figure A-10 Subsets of molecular formulae a – c) significantly more abundant in cluster 2 samples (see main text for definition of cluster 2), d – f) non-significantly related to cluster 1 or 2 (i.e., similar in ion abundance in samples of both clusters, thus "whole set"), and g – i) non-significantly related to blackwater-influenced samples from both sites (i.e., similar in ion abundance in blackwater samples from both sites, thus general "blackwater"). See specific set of formulae of cluster 1, Campina, and riparian Cuieiras samples in Figure 8. Panels a, d and f show the average and max mass spectra of each formula subset. Panels b, e and h show the formula subsets in Van Krevelen space (see formula classes and respective numbers of formulae in the legend in panels c, f and i). Panels c, f, and i show formula subsets in H/C vs. m/z space. Note the striking similarity of formula subsets related to samples of cluster 2 (a – c) and "blackwater" samples (g – i).
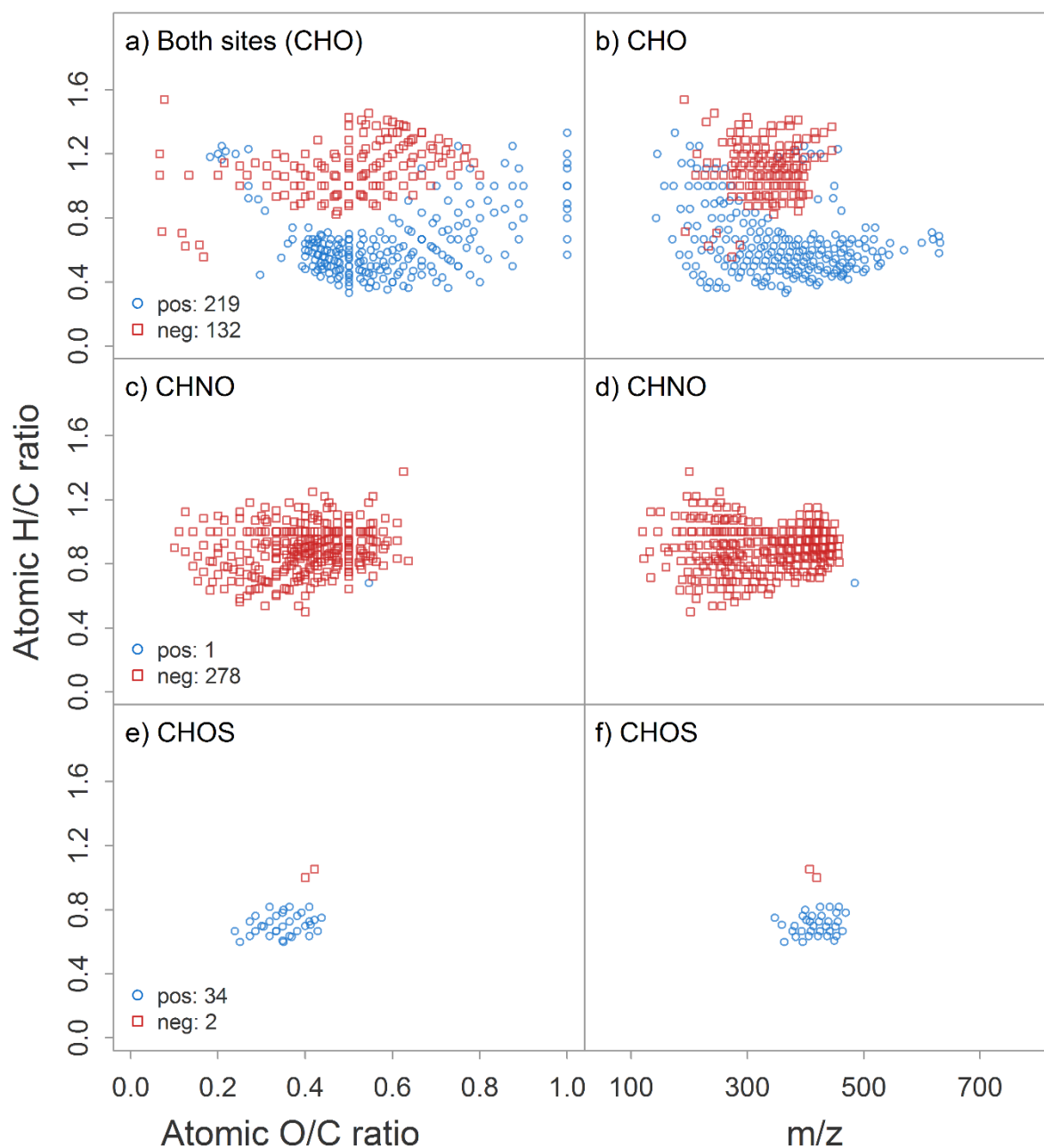
Figure A-11 Radiocarbon-correlated formulae in SPE-DOM exported from whitesand areas in central Amazonia, assessed for the whole sample set. Formula classes (CHO, CHNO, CHOS, rows) are shown in Van Krevelen (left panels) and H/C vs. m/z space (right panels). Positive correlation with F14C values relates to seemingly older DOM. The correlation pattern is strongly superposed by a site effect (Campina, blue, ^{14}C-older DOM; Cuieiras, red, ^{14}C-younger DOM) and thus required reanalysis for the two sites separately (Figure A-12; Figure A-13) a, b) CHO formulae separated well into two clusters. c, d) A large cluster of CHNO formulae is correlated negatively with F14C values. e, f) In contrast, a distinct cluster of CHOS formulae is correlated positively with F14C.
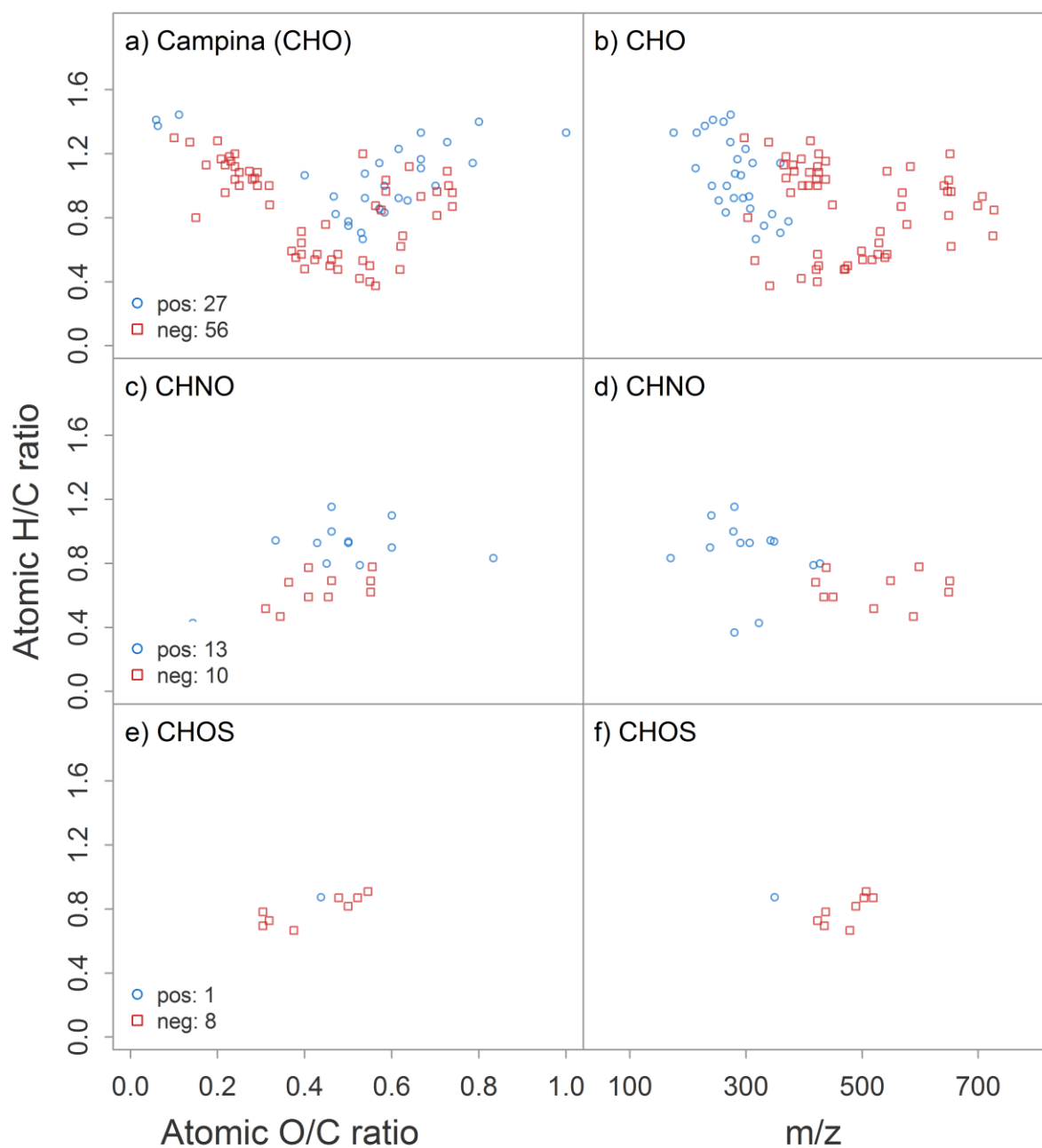
Figure A-12 Radiocarbon-correlated formulae in SPE-DOM exported from the Campina site. Positive correlation with F14C values relates to seemingly older DOM. Formula classes (CHO, CHNO, CHOS, rows) are shown in Van Krevelen (left panels) and H/C vs. m/z space (right panels). a, b) CHO formulae; c, d) CHNO formulae; e, f) CHOS formulae.
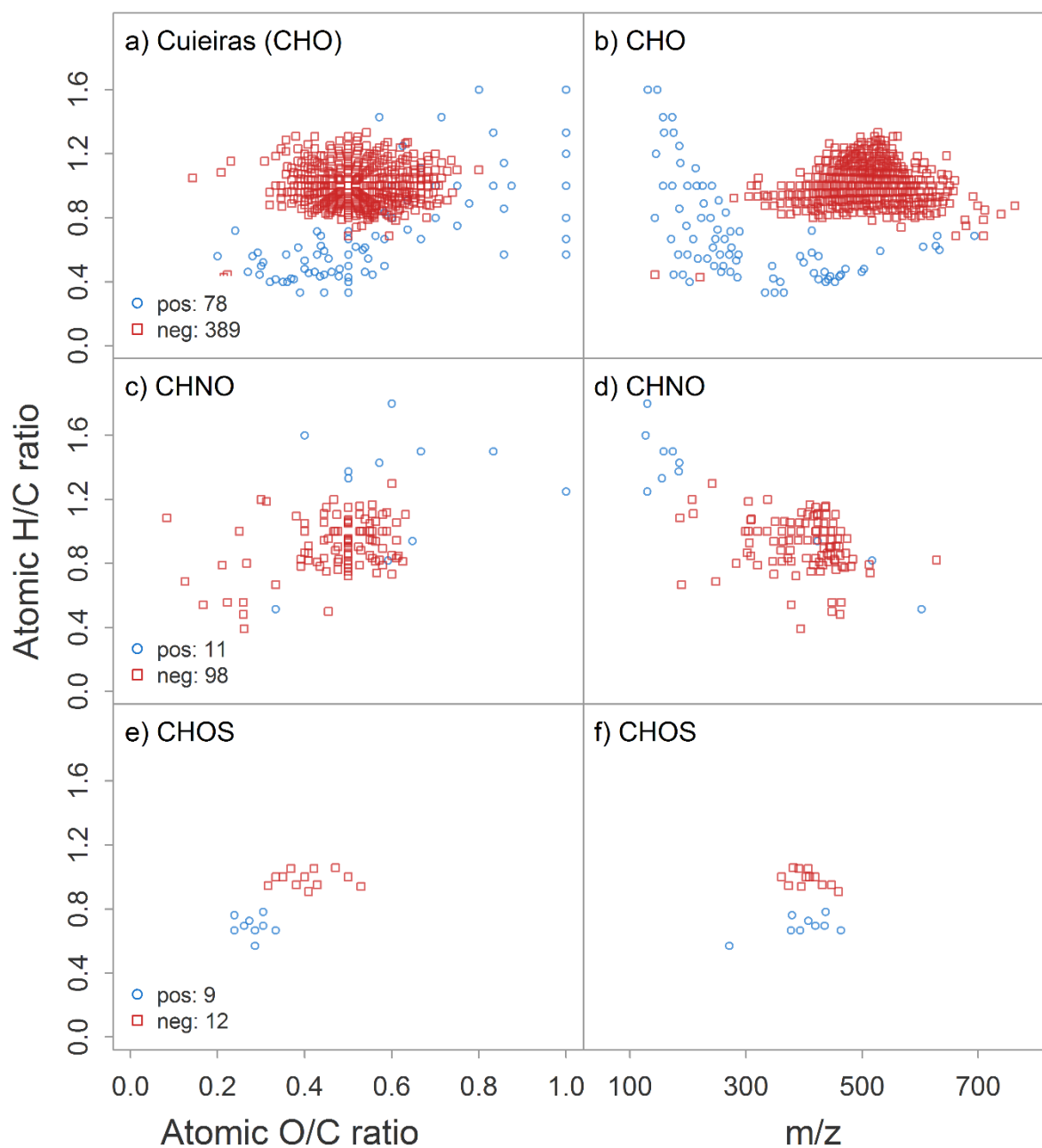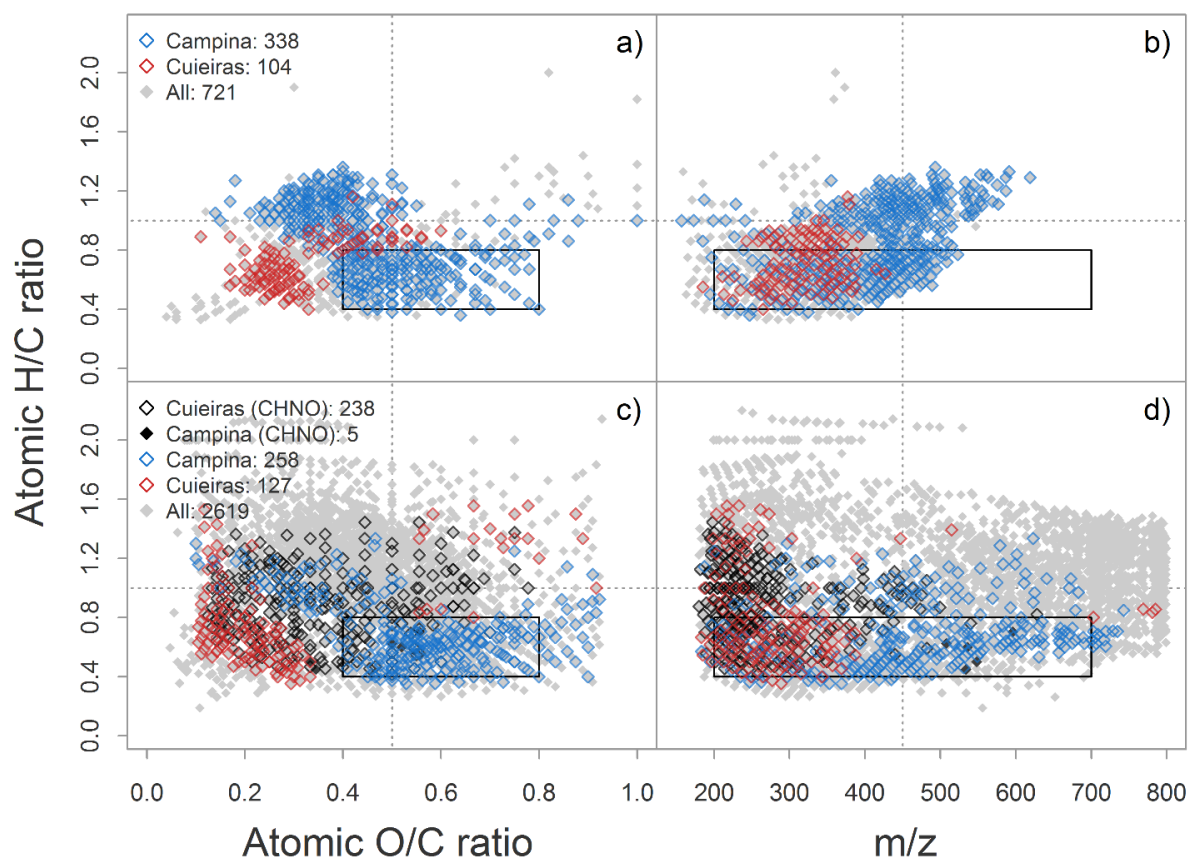
Figure A-13 Radiocarbon-correlated formulae in SPE-DOM exported from the Cuieiras site. Positive correlation with F14C values relates to seemingly older DOM. Formula classes (CHO, CHNO, CHOS, rows) are shown in Van Krevelen (left panels) and H/C vs. m/z space (right panels). a, b) CHO formulae; c, d) CHNO formulae; e, f) CHOS formulae.

Figure A-14 Overlap of site-specific formulae sets (see legend, CHO formulae in blue and red, CHNO formulae in black) with Rio Negro-specific formulae defined in a broad manner (see main text) as seen in Van Krevelen space (left panels) and H/C vs. m/z space (right panels). Small grey diamonds are total sets of Rio Negro-specific formulae. Rio Negro-specific formulae were extracted from a, b) Simon et al., (2019) and c, d) Gonsior et al., (2016). The black box denotes the area of most robust specific formulae in Van Krevelen space (H/C $0.4 - 0.8$; O/C $0.4 - 0.8$) and is also transferred to H/C vs. m/z space although not as specific (relatively wide mass range of specific signals, m/z $200 - 700$).
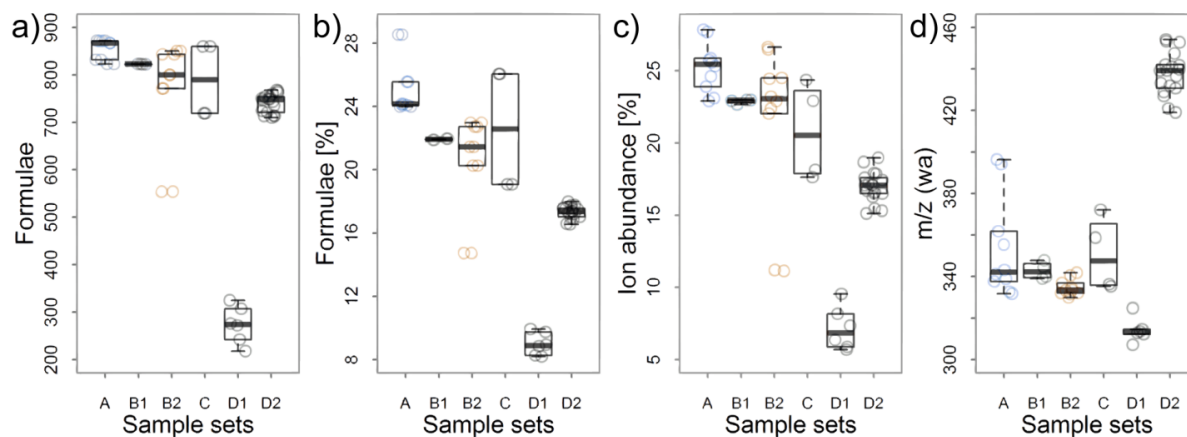
Figure A-15 Behavior of formulae in the core area of Rio Negro-specific signals (H/C 0.4 – 0.8; O/C 0.4 – 0.8), by number of formulae (a), percentage of formulae (b), relative ion abundance (c), and ion-abundance-weighted average m/z of formulae (d) across a hypothetical hydrological gradient of Campina piezometers (sample set A), Cuieiras lysimeters (B1), Cuieiras piezometers (B2), headwater streams (igarapés) of both whitesand sites (C), Rio Negro samples of the mainstem and smaller tributaries close to Manaus (D1, Simon et al., 2019), and Rio Negro mainstem and adjacent lakes close to Novo Airão, 120 km NW of Manaus (D2, Gonsior et al., 2016).
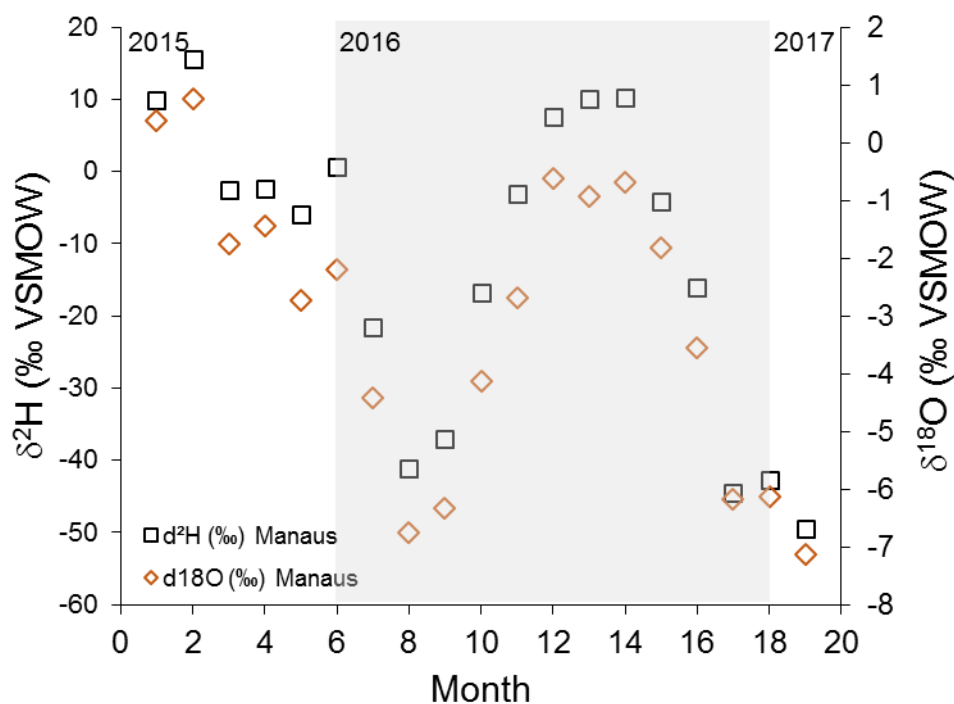
Figure A-16 Water isotope data from INPA's meteorological station in Reserva Ducke, north of Manaus, over the course of the year (data coverage: 08/15 – 02/17; n=19). Grey box denotes 2016; from 01/2016 – 01/2017. Water is isotopically light during the wet season (February - May), and becomes heavy in the dry season, peaking from July- September.
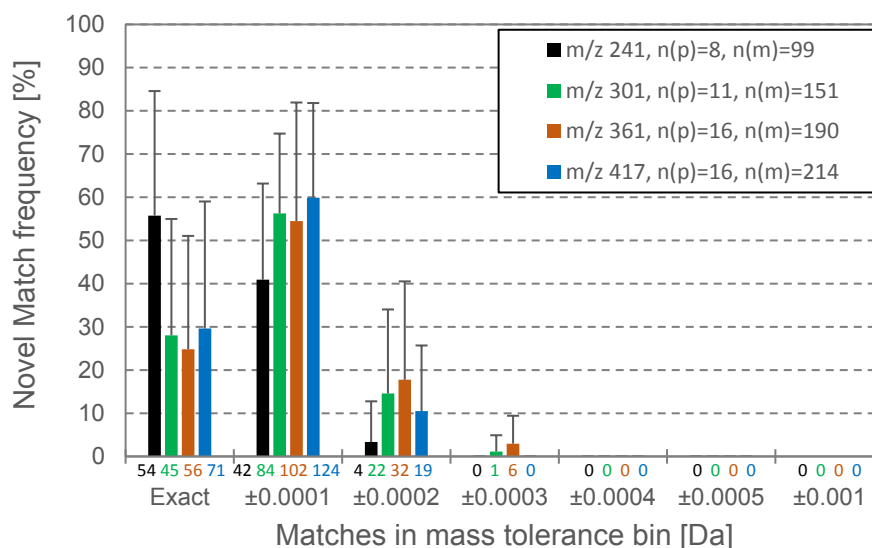
Figure A-17 Match frequency of non-indicative MDs (Table A-10) a mass range from 2 – 193 m/z. Data for each of the four IPIMs is presented (colors, see legend) along with total number of precursors "n(p)" and total number of matches across these precursors "n(m)". Small numbers below bars indicate absolute numbers of matches (average over all precursors of the respective IPIM). Error bars are +1SD across all precursors. Match frequency was then plotted vs. mass tolerance bin (x-axis), indicating how many % of matches were found in each bin, starting from the exact MD ("exact", exact mass to four digits). The tolerance bin was increasingly widened and the number of additional ("novel") matches – i.e., those not detected at narrower bin size – was monitored. The plot shows that the majority of matches to non-indicative MDs were found within the applied tolerance window (± 0.0002 Da). It also shows that outside of this window, the matching frequency drops immediately close to zero, indicating a low match rate in terms of detecting false positives, even when widening the tolerance bin to ±0.001 Da. Note, the analysis of each precursor ion also included a number of MDs showing no matches within the ± 0.0002 Da tolerance window (often the majority; however we only used precursors here that showed at least seven MD matches which translates into a maximum of 47 negative "hits", number of MDs in the non-indicative list = 54). Also for those MDs not matched within the applied tolerance window of ± 0.0002 Da, we found no novel (additional) matches in the widened tolerance bins (data included in the Figure) indicating that the MD approach is selective to losses that make chemical sense: We would expect random matches if the calculated MDs were derived from noise and not from an inherently structured biogeochemical signal. It also indicates that the peaks of interest are adequately resolved.
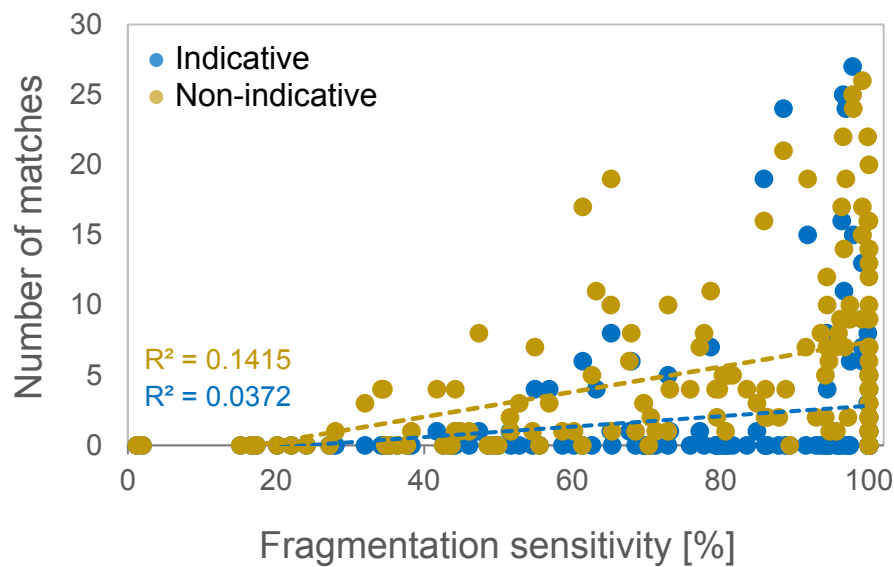
Figure A-18 Number of matches against lists of non-indicative MDs in DOM (ochre, Table A-10) and indicative MDs exclusively observed in standard compound data (blue, Table 5-1) in relation to the precursor's degradability (% change in ion abundance) between CID 0 (non-fragmented) and CID 25. Degradability is a poor predictor of match number but obviously a precursor needs to fragment to some degree in order to indicate positive matches. All precursors across the four IPIMs (n=159) are shown, and best fit-curves are linear regressions.
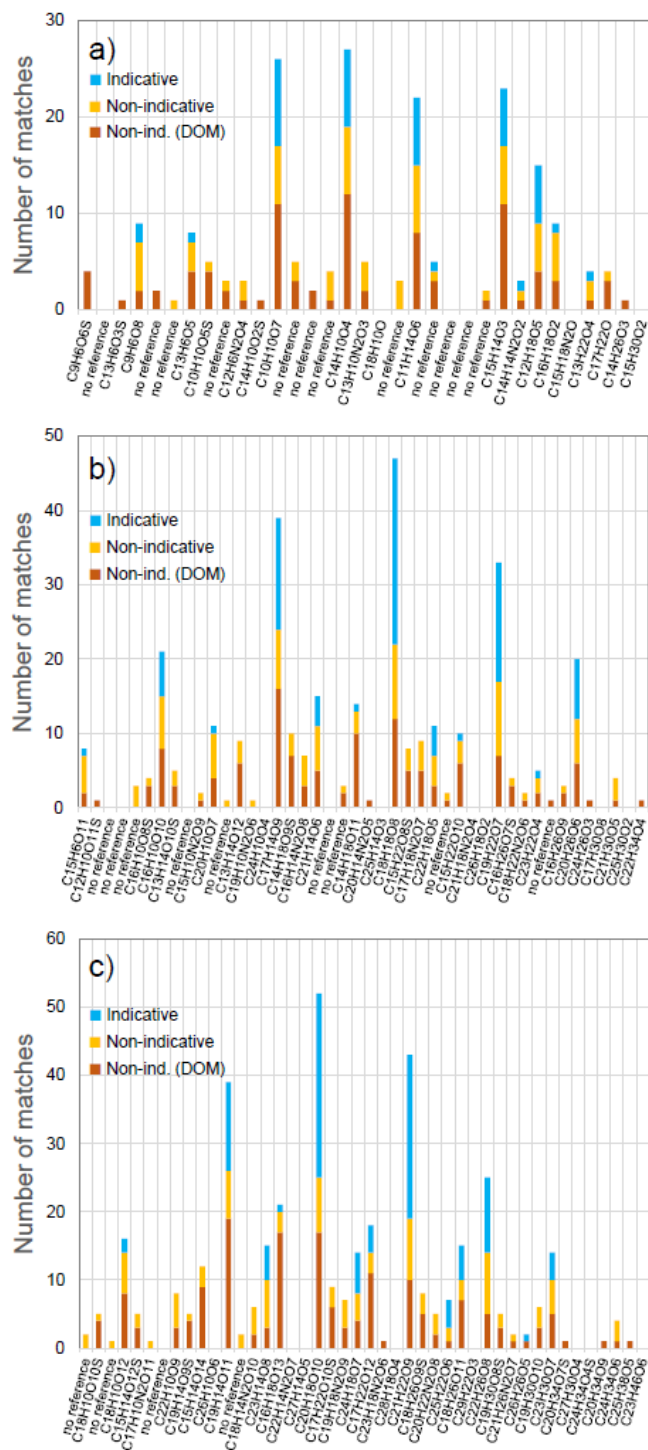
Figure A-19 Matching profiles of unknown precursors (molecular formulae given if assigned, otherwise, "no reference") for IPIMs at a) m/z 241, b) m/z 361, and c) m/z 417 (m/z 301, see Figure 5-3b in the main text). "Non-indicative" designates the list of reported MDs for DOM (Table A-10); "indicative" are those MDs found in exclusively in standard compounds in this study (Table 5-1); Table 5-2 in the main text shows those MDs that were part of both lists ("Non-indicative" in yellow color).
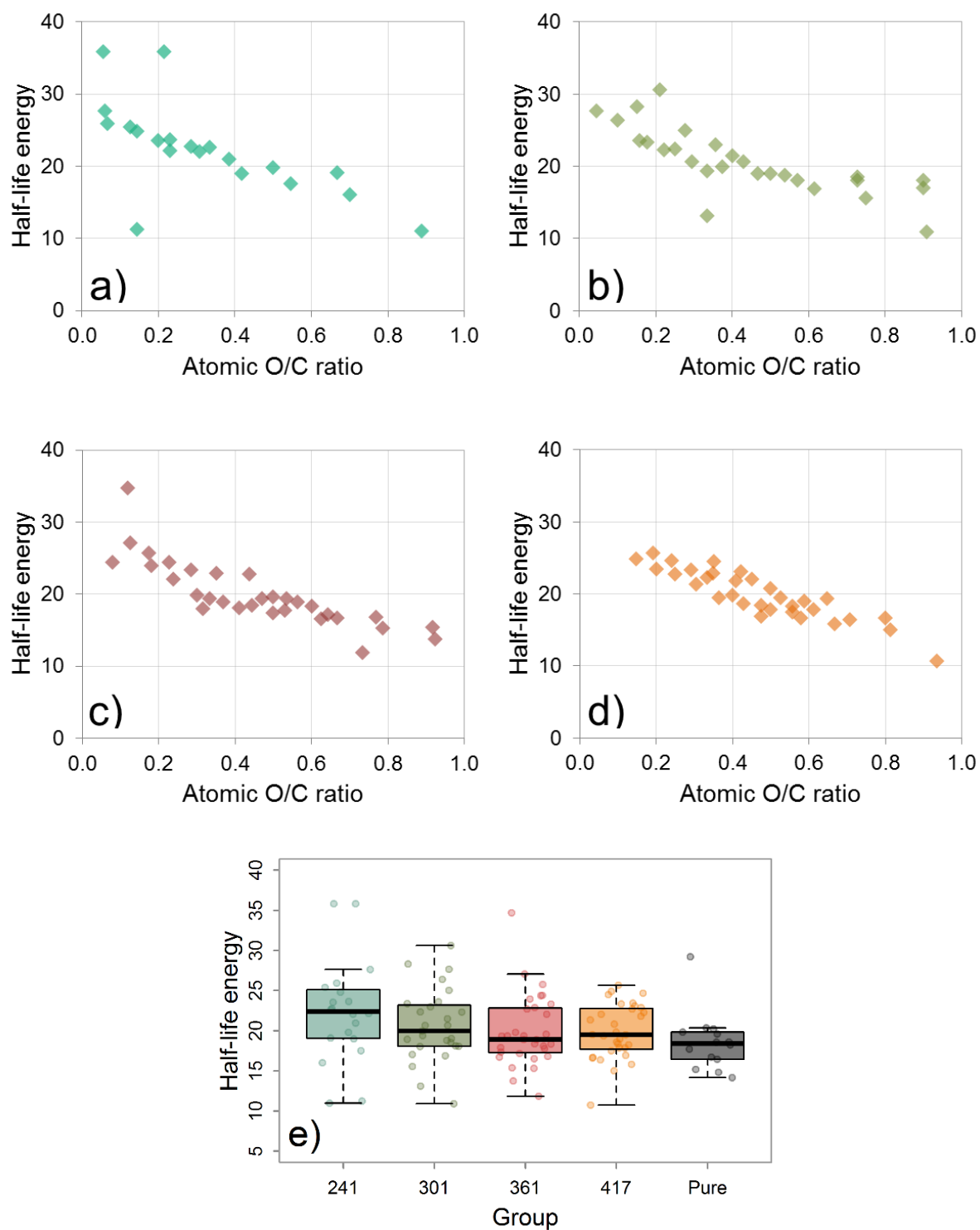
Figure A-20 Demonstrating the presence of a fragmentation sensitivity continuum in DOM precursor mixtures (IPIMs). a – d): Half-life energy plotted vs. O/C ratio of formulae at m/z 241, 301, 361 and 417 (same order). Half-life energy denotes the normalized collision energy at which initial ion abundance has decreased by 50%. High values indicate low fragmentation sensitivity. Panel e) Comparison of continua across nominal masses and for the standard compound set.
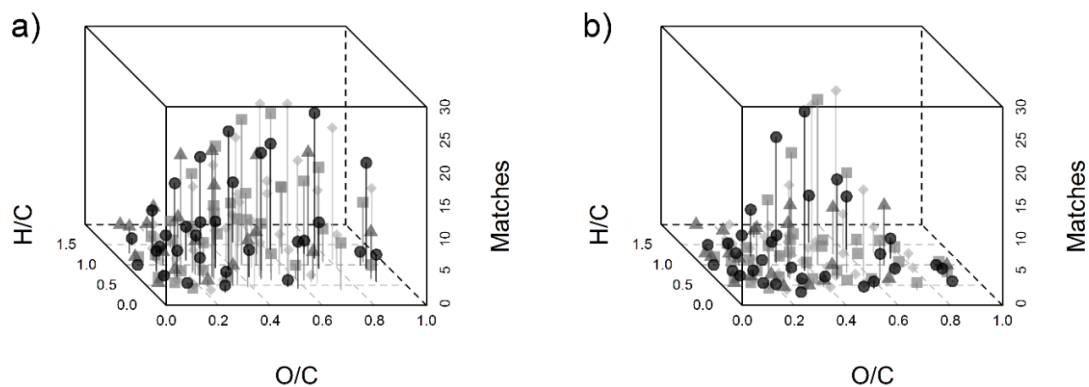
Figure A-21 3D-Van-Krevelen plots of the matching frequency to a) non-indicative MDs (Table A-10) and b) indicative MDs (Table 5-1). Matching patterns differ substantially, indicating that different factors determine the matching frequency of a precursor. In a), the oxygen content of a formula is a major controlling factor, besides initial intensity. In b), initial intensity is the most important factor. Also formulae with zero matches are shown. Legend: refer to Figure 5-4 in the main text.

Notes

Note A–1: Application of a method detection limit (MDL)

The used MDL approach (Riedel and Dittmar, 2014) approximates noise height by analyzing the intensity (i.e., ion abundance) distribution in the signal region between 0.3 and 0.9 Da for each nominal mass. DOM signal "islands" are usually stretched over a mass defect range from -0.1 Da to +0.3 Da. Noise detection can be implemented on the sample data itself that way, by this also saving instrument time. A problem encountered with the Orbitrap data were peaks around $0.5 - 0.65$ Da, being indicative of double-charged species. These signals were thus excluded for noise assessment. We used only those nominal masses that contained more than 20 noise signals, and only those showing a normal distribution (of intensity data, assessed by Student's t-test; Riedel & Dittmar, 2014). Noise peak ensembles showing p-values > 0.05 in the t-Test (rejecting the assumption of non-normality) also showed small deviation in a QQ-plot assessment (deviation from the normality line). By applying this MDL approach to the Orbitrap data, we found that the "20 peaks"- limit and the normality criterion were not fulfilled by all nominal masses, especially at higher m/z. This is partly due to decreasing resolution with m/z, but was also negatively influenced by the data acquisition mode. The default mode of data acquisition on the Orbitrap Elite ("reduced profile mode") influences mainly the noise peak yield per nominal mass. This mode of data acquisition reduces file size and scan time and is thus pivotal for high throughput data analysis with regular computing power (standard run with 100 scans, 22 kB file size in reduced vs. 800 kB in full mode; one standard scan, 1 sec in reduced vs. 2.2 sec in full mode). However, this form of acquisition by intention rules out the majority of the noise to improve signal detection, and thus hinders application of the MDL approach. Future studies need to take this into account and should determine the MDL based on full-profile data or at least compare reduced-profile data to it. Despite the named drawbacks, the MDL levels of nominal masses fulfilling the normality criterion were used here for an estimation of a linear regression over m/z and were extrapolated up to m/z 1000 for each sample individually. Each sample was corrected with its own method detection limit (MDL) approximation, set at a conservative confidence level of 99.8 %. For the samples under study, MDL levels increased with m/z, while they were rather constant and generally lower for blank samples with only ~ hundreds of signals.

Selbstständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und unter Verwendung der

angegebenen Hilfsmittel, persönlichen Mitteilungen und Quellen angefertigt habe.

_____

Jena, den   Unterschrift des Verfassers