

Molecular links between whitesand ecosystems and blackwater formation in the Rio Negro watershed

Carsten Simon^{1,1,1}, Tania Pena Pimentel^{2,2,2}, Maria Terezinha F. Monteiro^{2,2,2}, Luiz Antonio Candido^{3,3,3}, Didier Gastmans^{4,4,4}, Heike Geilmann^{1,1,1}, Regison da Costa Oliveira^{2,2,2}, João Batista Rocha^{2,2,2}, Elaine Pires^{2,2,2}, Carlos A. Quesada^{2,2,2}, Bruce Forsberg^{5,5,5}, Sávio J. Filgueiras Ferreira^{2,2,2}, Hillândia Brandão da Cunha^{2,2,2}, and Gerd Gleixner^{6,6,6}

¹Max Planck Institute for Biogeochemistry

²Instituto Nacional de Pesquisas da Amazônia

³Instituto Nacional de Pesquisas da Amazonia

⁴São Paulo State University

⁵instituto nacional de pesquisas da amazonia

⁶Unknown

November 30, 2022

Abstract

Tropical rivers constitute a major portion of the global aquatic C flux entering the ocean, and the Rio Negro is one of the largest single C exporters with a particularly high export of terrestrial C. We investigated the role of whitesand ecosystems (WSEs) in blackwater formation in the Rio Negro basin to develop novel constraints for the terrestrial carbon export from land to the aquatic continuum. To this end, we used ultrahigh resolution mass spectrometry (FT-MS, Orbitrap) to identify markers in dissolved organic carbon (DOC) from ground- and surface waters of two contrasting WSEs feeding Rio Negro tributaries, and compared them with known Rio Negro marker from two openly available FT-MS datasets. Tributaries were fed by a whitesand riparian valley connected to terra firme plateau, and a typical upland whitesand Campina. WSE-DOC molecular composition differed by 80% from plateau DOC, which was characterized by reworked, highly unsaturated N- and S-containing molecules. WSE-DOC contained mainly condensed aromatics and polyphenols. WSE samples differed by 10% in molecular DOC composition and also by their isotopic content (14C, 18O, 2H). Upland WSE-DOC was exported by fresh precipitation and had maximum age of 13 years, being five years older than riparian valley WSE-DOC. Unexpectedly, only markers from the upland WSE, which cover a small proportion of the landscape, were identical to Negro markers. Markers of the riparian valley WSE, which are widespread and known for high DOC export, surprisingly showed lower coverage with Negro markers. Analysis of robust matching WSE markers between FT-MS datasets by Pubchem suggested well-known plant metabolites (chromenes and benzofurans) as promising candidates for targeted approaches and calibration. Our results suggest that terrestrial DOC from upland WSEs is a main source of specific blackwater molecules missing in the regional ecosystem C balance, whereas C export from the riparian valley and especially from terra firme plateaus represents mainly recycled and transformed carbon not directly affecting the ecosystem C balance. Our study highlights the potential of high-resolution techniques to constrain carbon balances of ecosystems and landscapes. Comparisons of FT-MS datasets and complementary isotopic information shows high potential to identify robust molecular markers that link forests, soils, aquifers and aquatic systems, and are needed for a deeper understanding of the regional C cycle in tropical blackwater catchments.

Molecular links between whitesand ecosystems and blackwater formation in the Rio Negro watershed

C. Simon^{(1),(2)}, T. P. Pimentel⁽³⁾, M. T. F. Monteiro⁽³⁾, L. A. Candido⁽³⁾, D. Gastmans⁽⁴⁾, H.

Geilmann⁽⁵⁾, R. da Costa Oliveira⁽³⁾, J. B. Rocha⁽³⁾, E. Pires⁽³⁾, C. A. Quesada⁽³⁾, B. R.

Forsberg^{(3),(6)}, S. J. F. Feirrera⁽³⁾, H. B. da Cunha⁽³⁾, G. Gleixner^{*(1)}

⁽¹⁾ Molecular Biogeochemistry, Max Planck Institute for Biogeochemistry (MPI-BGC), Hans Knöll-Str. 10, 07745 Jena, Germany.

⁽²⁾ present address: Institute of Biogeochemistry and Pollutant Dynamics (IBP), ETH Zürich, Universitätstrasse 16, 8092 Zürich, Switzerland.

⁽³⁾ Coordenação de Dinâmica Ambiental (CODAM), Instituto Nacional de Pesquisas da Amazônia (INPA), Av. Efigênio Sales 2239, Aleixo, Manaus, Brazil.

⁽⁴⁾ São Paulo State University (UNESP), Centro de Estudos Ambientais, Av. 24A, 1515, Bela Vista, Rio Claro, São Paulo, Brazil.

⁽⁵⁾ Stable Isotope Laboratory (BGC-IsoLab), Max Planck Institute for Biogeochemistry, Hans Knöll-Str. 10, 07745 Jena, Germany.

⁽⁶⁾ present address: Vermont Agricultural and Environmental Laboratory, 163 Admin Dr, Randolph Center, 05602, Vermont, United States.

*Correspondence: gerd.gleixner@bgc-jena.mpg.de (Gerd Gleixner)

Highlights

- Tropical blackwater catchments are a hotspot of aquatic C export
- Exported C represents a major portion of NEP, and stores molecular information
- Robust Rio Negro basin ecosystem markers identified by molecular and isotopic tools
- Ecosystem markers linked to specific upland site but not widespread riparian system
- If properly calibrated, markers may serve as proxies of NEP or terrestrial C export

Abstract

24 Tropical rivers constitute a major portion of the global aquatic C flux entering the ocean, and the
25 Rio Negro is one of the largest single C exporters with a particularly high export of terrestrial C.
26 We investigated the role of whitesand ecosystems (WSEs) in blackwater formation in the Rio
27 Negro basin to develop novel constraints for the terrestrial carbon export from land to the aquatic
28 continuum. To this end, we used ultrahigh resolution mass spectrometry (FT-MS, Orbitrap) to
29 identify markers in dissolved organic carbon (DOC) from ground- and surface waters of two
30 contrasting WSEs feeding Rio Negro tributaries, and compared them with known Rio Negro
31 marker from two openly available FT-MS datasets. Tributaries were fed by a whitesand riparian
32 valley connected to *terra firme* plateau, and a typical upland whitesand *Campina*. WSE-DOC
33 molecular composition differed by 80% from plateau DOC, which was characterized by
34 reworked, highly unsaturated N- and S-containing molecules. WSE-DOC contained mainly
35 condensed aromatics and polyphenols. WSE samples differed by 10% in molecular DOC
36 composition and also by their isotopic content (^{14}C , ^{18}O , ^2H). Upland WSE-DOC was exported
37 by fresh precipitation and had maximum age of 13 years, being five years older than riparian
38 valley WSE-DOC. Unexpectedly, only markers from the upland WSE, which cover a small
39 proportion of the landscape, were identical to Negro markers. Markers of the riparian valley
40 WSE, which are widespread and known for high DOC export, surprisingly showed lower
41 coverage with Negro markers. Analysis of robust matching WSE markers between FT-MS
42 datasets by Pubchem suggested well-known plant metabolites (chromenes and benzofurans) as
43 promising candidates for targeted approaches and calibration. Our results suggest that terrestrial
44 DOC from upland WSEs is a main source of specific blackwater molecules missing in the
45 regional ecosystem C balance, whereas C export from the riparian valley and especially from
46 *terra firme* plateaus represents mainly recycled and transformed carbon not directly affecting the

ecosystem C balance. Our study highlights the potential of high-resolution techniques to constrain carbon balances of ecosystems and landscapes. Comparisons of FT-MS datasets and complementary isotopic information shows high potential to identify robust molecular markers that link forests, soils, aquifers and aquatic systems, and are needed for a deeper understanding of the regional C cycle in tropical blackwater catchments.

Keywords: Dissolved organic matter, critical zone, hydrochemistry, chemodiversity, soil organic matter, biomarker, land-to-ocean continuum, watershed, river basin

1 Introduction

The riverine export of terrestrial dissolved organic carbon (DOC) constitutes a major flux within the boundless carbon cycle that connects land and ocean (Regnier et al., 2013; Drake et al., 2018b; Webb et al., 2018). Global estimates indicate a total of 208 ± 28 Tg DOC exported by rivers each year (Dai et al., 2012), mainly coming from large tropical and circumboreal watersheds (Raymond and Spencer, 2014; Li et al., 2019). Tropical rivers account for a major portion of the flux, being equivalent to 62-66 % of global DOC export (Dai et al., 2012; Huang et al., 2012). Moreover, recent modelling efforts show that tropical annual exports have been rising over the past 65 years, with an increase of 10 Tg C compared to 1960 in case of South America (Li et al., 2019). The three tropical rivers with highest discharge, the Amazon (incl. Tocantins), the Congo and the Orinoco, alone deliver 18% of global riverine DOC (Raymond and Spencer, 2014). Due to within-river DOC transformations, actual export of terrestrial organic carbon expected to be even higher (Drake et al., 2018b), and novel markers are needed to quantify the original terrestrial part of the carbon export.

Blackwater river basins stand out as hotspots of DOC release in the tropics (Junk et al., 2011). The Rio Negro basin for example, which covers roughly 10% of the area of the Amazon river basin, accounts for an annual DOC export of 5.2 – 6.7 Tg C (Coynel et al., 2005; Guinoiseau et al., 2016) equivalent to 17-23% of the Amazon's total DOC export (Raymond and Spencer, 2014). Tropical blackwater rivers such as the Rio Negro are thus pivotal in understanding the global carbon cycle and its response to environmental change (Alvarez-Cobelas et al., 2012; Raymond and Spencer, 2014; Webb et al., 2018).

Qualitative analyses of DOC composition have shown great potential to track processes and source contributions on the watershed scale (Creed et al., 2015; Riedel et al., 2016; Hutchins et al., 2017; Drake et al., 2019; Spencer et al., 2019). Processes such as deforestation (Drake et al., 2019; James et al., 2019), drainage (Moore et al., 2013) or warming (Drake et al., 2018a) all affect the molecular composition or age of leached DOC. Progress in qualitative DOC measures could thus effectively complement global and regional modelling efforts based on quantitative DOC export data (Alvarez-Cobelas et al., 2012; Lv et al., 2019). However, we do know surprisingly little about potential molecular markers in DOM and their environmental fate. Traditional targeted approaches aiming at terrestrial source markers, namely lignin phenols, can be questioned due to findings suggesting their fast turnover in soils (Gleixner et al., 2002; Hernes et al., 2007; Marschner et al., 2008) and within the land-to-ocean continuum (Cao et al., 2018), and potential autochthonous sources in marine systems (Powers et al., 2019). Only a few authors have addressed the traceability and stability of ecosystem imprints within the aquatic land-to-ocean continuum, and its conditions, in a non-targeted way by means of ultrahigh resolution mass spectrometry to identify novel markers (FT-MS; Roth et al., 2014; Medeiros et al., 2016; Hutchins et al., 2017; Wagner et al., 2019). This gap is due to the limited availability of these

molecular-level analytical tools. FT-MS techniques allow unprecedented molecular insight by resolving thousands of signals within a single DOM sample, which are assigned molecular formulae based on exact mass (Hertkorn et al., 2013). FT-MS techniques have now added important detail in character and transformation of ecosystem imprints at all stages of the aquatic continuum (Hutchins et al., 2017; Raeke et al., 2017; Kellerman et al., 2018; Lynch et al., 2019; Roth et al., 2019; Wagner et al., 2019). Dedicated sets of novel, traceable ecosystem markers are however rare and need to be calibrated to complement high-resolution DOC flux data (Roth et al., 2014; Medeiros et al., 2016; Cao et al., 2018). Robust sets of molecular markers could promote better understanding of ecosystem-resolved DOM export dynamics and its drivers. This knowledge is pivotal to evaluate and predict the vulnerability and biogeochemical functionality of watersheds under environmental change scenarios (McGuire et al., 2014; Abbott et al., 2018; Bernhardt et al., 2018; Jehn et al., 2020).

As described above, the Rio Negro basin is one of the world's largest DOC emitters in terms of estimated annual flux (6.7 Tg DOC) and yield (9.7 g DOC m⁻²), making it a classic “blackwater” river (Coynel et al., 2005; Dai et al., 2012). Scientists early noted the co-occurrence of tropical whitesand ecosystems (WSEs) and blackwater streams, and hypothesized a link between them (Sioli, 1954; Janzen, 1974; Leenheer, 1980; Goulding et al., 1988; Junk et al., 2011). The most characteristic feature of WSEs in these landscapes is their sandy soil, classified as either podzol (2% of Amazon basin area) or arenosol (3%) by the World Reference Base (Quesada et al., 2011). These soils differ largely from the more widespread clayey tropical soils such as ferralsols (32%), acrisols (29%), or plinthosols (9%) that are typically found on plateaus and their slopes (Do Nascimento et al., 2004; Quesada et al., 2011; Lucas et al., 2012). Due to the low water-holding capacity of sand, WSEs are also characterized by specific types of

vegetation that differ from highly diverse terra firme rainforests which are dominated by large trees. In the central Amazon, these are sclerophyllous, shrubby and smaller-tree dominated Caatinga, Campina, and Campinarana forests, known for their unique plant secondary metabolites and adapted decomposer communities (Janzen, 1974; Klinge and Medina, 1979; Zanchi et al., 2015; Demarchi et al., 2018; Vasco-Palacios et al., 2018). WSEs occur as local upland depressions on plateaus, as intersected valleys forming large riparian corridors at the foot of plateau slopes, or in low-elevation terrain in the form of wide peneplains (Do Nascimento et al., 2004; Montes et al., 2011). Roughly, upland and riparian valley WSEs are dominant in the lower Rio Negro basin while peneplains are widely distributed in the upper Rio Negro basin (Montes et al., 2011). It is likely that the molecular composition of DOM resolves the different environmental conditions, allowing for proper landscape-based DOC source identification and export calculations.

Previous studies that assessed terrestrial sources of Rio Negro DOM largely supported the older hypotheses that WSEs, and more specifically the widely distributed riparian corridors are responsible for the highest amount of carbon export (Junk, 1993; Remington et al., 2007; Melack and Hess, 2010; Bardy et al., 2011). However, this DOC export is mainly controlled by precipitation amount and flooding events (McClain et al., 1997; Remington et al., 2007; Zanchi et al., 2015). This is in line with the finding that in riparian settings, DOC generally shows transport-limited behavior, meaning that its absolute flux scales with discharge (Musolff et al., 2017; Zarnetske et al., 2018). As a result, water passing through the riparian zone continuously leaches existing reserves of processed organic matter (Ledesma et al., 2015; Laudon and Sponseller, 2018; Tiegs et al., 2019). In line with the older hypothesis, McClain and coworkers reported low annual DOC yields for widespread plateau areas (terra firme – ferralsol; 2 g DOC

m-2 yr-1) but large yields for a relatively small Campina WSE catchment (40 g DOC m-2 yr-1) in the lower Negro basin, north of Manaus. The authors predicted that a WSE molecular DOM signal would thus be easily detectable in higher order rivers within the Rio Negro catchment (McClain et al., 1997), and later studies conducted in the same region corroborated this hypothesis (Remington et al., 2007). DOM from well-developed podzols reflected best the chemical properties of DOM isolates from local groundwater and nearby rivers, showing the fast transit of DOM in well-drained sandy soils with low sorption potential (Remington et al., 2007; Bardy et al., 2011). However, no subsequent markers were identified or tracked in the lower reaches of the stream network to assess their environmental fate or marker potential (Bardy et al., 2011). Such novel markers are however needed to study variations in DOM export and spatiotemporal dynamics in riverine DOM sources within a catchment (Hutchins et al., 2017; Bernhardt et al., 2018; Laudon and Sponseller, 2018).

Recently, watershed-specific molecular DOM signatures of the Rio Negro and other Amazon tributaries (Tapajos, Madeira, Solimões) were reported (Gonsior et al., 2016; Simon et al., 2019) that could serve as potential markers of ecosystem DOC exports due to their largely conservative behavior during mixing (Simon et al., 2019). We here make use of these openly available FT-MS DOM datasets and compare them to groundwater, surface, and soil water DOM measured by Orbitrap FT-MS. We hypothesized that the overall large export of DOM from riparian WSEs in the Rio Negro basin would allow for the retrieval of Rio Negro-specific markers as assumed by previous studies (McClain et al., 1997; Remington et al., 2007; Bardy et al., 2011). We therefore compared an upland Campina forest, and a riparian valley system dominated by Campinarana forest, both typical for WSE-podzol systems within elevated terra firme plateaus north of Manaus. We hypothesized that both WSEs and plateaus would differ in terms of water chemistry

and DOC properties, and that the DOM molecular composition would reflect these differences as well, yielding new sets of unique ecosystem markers. We further hypothesized that riparian valley WSE markers would indicate clear overlap with known Rio Negro markers, and could thus serve as complementary proxies of land-derived primary production in the Rio Negro basin.

2 Materials and Methods

2.1 Field sites and sampling procedures

Soil water samples were taken in early November 2017 at the onset of the rainy season in two protected forest reserves under the responsibility of the Instituto Nacional de Pesquisas da Amazônia (INPA) in Manaus, Brazil (Figure 1; Supporting Information Figure S1). 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 (Zanchi et al., 2014; Marques et al., 2016). We conducted sampling from 31st October – 2nd of November, and all lab procedures followed within three days. The geological setting, landscape structure, forest composition, and soil characteristics have been described in detail by Zanchi et al. (2014). Broad swampy valleys surrounded by elevated plateaus cover about half of the Cuieiras reserve's area (in total, 22735 ha) (Zanchi et al., 2014). The valley soils (podzols, gleysols) differ markedly from the clayey plateau and slope soils (oxisols, ultisols). Bleached quartz sand and high amounts of phenolics characterize the valley's podzols (Monteiro et al., 2014; Zanchi et al., 2014; Marques et al., 2016), and the presence of *Mauritia flexuosa* (Luizão et al., 2004), a palm species indicative of hydromorphic conditions, suggests poor drainage (Junk, 1993). A second-order blackwater stream (Rio Açu) drains the area, meeting with Rio Cuieiras, Rio Branquinho, and Rio Negro downstream (Monteiro et al., 2014). The catchment has an area of 660 ha (Monteiro et al., 2014).

Monteiro and coworkers report a total annual rainfall of 2806 mm in 2002 and 2004 mm in 2003. Stream discharge was 0.08 – 5.59 m³ s⁻¹ (average, 0.18 m³ s⁻¹) and DOC levels ranged from 3.2 – 15.2 mg C l⁻¹ (average, 8.7 ± 3.0 mg C l⁻¹) during that period. The watershed's annual total carbon flux was estimated as 13.3 g C m⁻² in 2003 (Waterloo et al., 2006) and Monteiro and coworkers estimated a stream DOC flux (Rio Açu) of 8.7 g C m⁻² for the same year. Zanchi and coworkers estimated the annual DOC export of the Açu watershed to lie within the range 9.3 – 22.7 g C m⁻² (Zanchi et al., 2015). Water samples were taken from piezometers installed across a valley transect (Monteiro et al., 2014) which is maintained and sampled regularly. Piezometers were emptied once before final sampling. The stream was sampled manually, with nitrile gloves, against the direction of flow, using pre-cleaned (acidified ultrapure water, pH2, HCl, Merck EMSURE®, p.a., ACS grade) Nalgene™ polycarbonate bottles (Fisher Scientific, Schwerte, Germany) that were cleaned with the respective sample before final sampling. We also sampled two deep wells on the plateau (35 m and 39 m depth) by lowering an empty, clean sampling bottle on a string until water was reached.

The Reserva Campina is a 900 ha reserve that shows only small relief; poor sandy soils (up to 99% sand) that co-occur with typical but specific forest types, so-called Campina and Campinarana forests (heath forests; Demarchi et al., 2018) characterize the area. In contrast to the highly diverse plateau (terra firme) forests, Campinas show much lower species diversity and low canopy (~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 headwater stream that is less than 1m wide and often less than 30 cm deep (McClain et al., 1997) and feeds the Rio Tãruma Açu in the southward direction, meeting with the Rio Negro close by Manaus (Figure 1). The catchment area is estimated by 6.5 ha and thus ~100 times smaller than the Rio Açu catchment

(Zanchi et al., 2015). According to the same authors, annual rainfall levels are comparable among both reserves. Surface runoff (from the stream) and estimated groundwater outflow amounted to 485 and 1071 mm in that period, respectively (in sum 1556 mm; as compared to Açu 1362 mm; Waterloo et al., 2006). Zanchi and coworkers estimated the total watershed's DOC export with 49.2 g C m⁻², including groundwater outflow and rainfall, of which the stream exported 15.3 g C m⁻². An annual streamflow of 485 mm equals an estimated discharge of 1.0 x 10⁻³ m³ s⁻¹, which is about 180 times lower than average discharge at Rio Açu in 2002 – 2003 (Monteiro et al., 2014). McClain and coworkers reported annual DOC exports of 40 g C m⁻² yr (McClain et al., 1997; Zanchi et al., 2015) from the same catchment in the period 1993 – 1994. We took samples at the side slopes of the stream from piezometers installed in 1993, as detailed in McClain et al. (1997). For this purpose, wells were emptied three times and sampled afterward (Zanchi et al., 2015). The stream was sampled as described above.

2.2 Water chemistry: TOC, pH, electrical conductivity

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-VCPH model, Shimadzu, Kyoto, Japan) (Monteiro et al., 2014). Before extraction, we analyzed samples for pH and electrical conductivity (EC) with a Multi 340i probe system (WTW, Weilheim, Germany).

2.3 Solid-phase extraction of DOM

DOM samples were solid-phase extracted (SPE) shortly after sampling at 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 (pH 2, HCl), and ultrapure methanol (Biotec Reagentes Analíticos, p.a., ACS grade). We loaded columns with maximal amounts of 3 mg C. The extraction efficiency (EE) of samples with high DOC concentrations ($> 2 \text{ mg l}^{-1}$, $n = 11$) was always $> 60\%$ and on average $71 \pm 8\%$, at a loading ratio of 443 ± 143 (average $\pm$ standard deviation; PPL: DOC in mg/ mg). Sample PT6 was an exception (high DOC, low EE: 33%). Samples with low DOC concentrations ($< 2 \text{ mg l}^{-1}$, $n = 4$) showed lower extraction efficiency ($23 \pm 15\%$), and loading ratios were generally higher (2400 ± 960) (Supporting Information Table S1).

2.4 Water isotopes

Isotopic signatures of water ($\delta^2\text{H}$ - and $\delta^{18}\text{O}$ -values) 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 the literature (Gehre et al., 2004). In short, 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 $\delta^2\text{H}$ and $\delta^{18}\text{O}$ values are reported on the VSMOW 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 $\delta^2\text{H}$ and $\delta^{18}\text{O}$ measurements are usually better than 1 and 0.1 ‰, respectively. We accessed regional isotope ratio estimates of rain for October and November 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 $\delta^2\text{H}$, and -2.4‰ and -3.8‰ (V-SMOW) for $\delta^{18}\text{O}$ in October and November, respectively. We estimated the average between both monthly values for each isotope, considering our sampling date at the end of October/beginning of November.

2.5 Radiocarbon analysis of solid-phase extracted DOC

We transferred an aliquot of methanolic PPL extract equivalent to 0.25 mg C to tin capsules (8 mm diameter, 20 mm height; IVA Analysentechnik, Meerbusch, Germany). The methanol was left to evaporate. Air-dried capsules were combusted in an elemental analyzer and graphitized for radiocarbon analysis on a 3 MV Tandetron 14C-AMS (HVEE, Amersfoort, Netherlands) at the Max Planck Institute for Biogeochemistry in Jena, Germany. Modern (Oxalic Acid II) and ^{14}C -depleted standard materials were carried along for quality control and data corrections (Steinhof et al., 2017; Benk et al., 2018). 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 ($F^{14}\text{C}$), which is the fraction of the standard concentration normalized for $\delta^{13}\text{C}$ of the oxalic acid standard measured at the same time (with $\delta^{13}\text{C}$ of -19‰; Trumbore et al., 2016). $\Delta^{14}\text{C}$ values also take into account radioactive decay of the oxalic acid standard since 1950, which demarks the start of bomb tests that increased the ^{14}C content of the atmosphere (Trumbore et al., 2016). We calibrated our background-corrected $F^{14}\text{C}$ data with atmospheric data published in Graven (2015) to derive DOC age. We used values for the inner-tropical convergence zone (zone SH3). To cover the time point of sampling (2017), we predicted $\Delta^{14}\text{C}$ values for 2016 and 2017 based on data from 1997-2015 by linear regression ($R^2 = 0.996$) and then estimated DOC mean

calendar age. The measurement error was < 1 year and < 2 years for repeated measurements (standard deviation).

2.6 Orbitrap measurements of DOM and data processing

We conducted Orbitrap measurements at the Max Planck Institute for Biogeochemistry, as described elsewhere (Simon et al., 2018), at a nominal resolution setting of 480,000. Allowable numbers of atoms in formula calculation were as follows: ^{12}C , 0-60; ^1H , 0-120; ^{14}N , 0-2; ^{32}S , 0-1; ^{16}O , 1-60, ^{13}C , 0-1. The assignment was done at ± 1 ppm tolerance. We defined noise as the smallest peak in the sample set, not including zeros. We then calculated the maximal signal-to-noise ratio (S/N_{max}) of a peak across all samples (without blanks and reference material) and only kept peaks with a value > 5. We calculated the number of matches of each peak across all samples (without blanks and 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 (S/N_{max} < 20). After the exclusion of peaks outside the usual mass defect range of natural organic compounds (nominal m/z in Da, -0.05 mDa ... +0.3 mDa), 13658 peaks remained in the dataset in total. Of those, 9893 had a formula assigned, and 3672 included 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 N₂S were discarded due to their unlikely occurrence in DOM (Hawkes et al., 2016). In the case of ambiguous peaks with more than one assigned molecular formula, only suggested CHO formulae with a DBE-O of < 10 (modulus) were kept (Herzprung et al., 2014). 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 (only ^{12}C) formula were excluded. 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, we normalized all mass spectra to the sum of their peak intensities (including all peaks $> \text{S/N} = 5$, also those with no assigned formula). The further analysis of the data focused on the subset of peaks with an assigned monoisotopic formula ($n = 5709$). The crosstab is available from <https://doi.org/10.1594/PANGAEA.922606> (Supporting Information Data Set S1).

2.7 Statistical analyses: Ecosystem fingerprints and molecular links

We analyzed the molecular formula data 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` (Oksanen, 2010). 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 (Osterholz et al., 2016). We then analyzed the distribution of samples in coordinate space for clustering (indicating similarity among samples) and correlations with specific DOM indices by the `envfit` function (at 999 permutations). DOM indices aggregate properties of the molecular formula population of each sample (Supporting Information Table S2). We further assessed molecular formulae with a significant ($p = 0.05$) association to clusters of samples derived from PCoA analyses by Student's t-test of averaged relative ion abundances across samples of each group. We conducted two-sided tests assuming unequal variances.

To link headwaters and downstream signals, we compared sets of specific molecular formulae to known Rio Negro-specific markers available from two open-access FT-MS datasets (Gonsior et al., 2016; Simon et al., 2019). The dataset from 2016 compared samples from the Rio Negro and its adjacent lakes (close to Novo Airão, 120 km northwest of Manaus) to distant large rivers (Rio Tapajos, Rio Madeira) to reveal large-scale differences in chemodiversity, i.e., unique signals of each river basin. We extracted the robust Rio Negro fingerprint by selecting unique formulae of the Rio Negro detected across all measurements, with an average ion abundance of at least 5×10^7 (or $\sim 0.01\%$ relative intensity; $n = 225$; Gonsior et al., 2016). The dataset published in 2019 assessed differences of riverine DOM at the confluence of the Amazon near Manaus (Encontro das Águas). To obtain a robust DOM fingerprint, we extracted formulae that showed a significant positive correlation (Pearson's r , $p = 0.05$) to the fraction of Rio Negro during mixing in each of the studies' three experiments (Simon et al., 2019; $n=299$). The datasets were also compared on a general level. Information on this aspect is presented in the Supporting Information (Supporting Information Text S1). The combined data from all three studies is available in presence/absence format and with ion abundance information from <https://doi.org/10.1594/PANGAEA.922606> (Supporting Information Data Set S2).

3 Results

3.1 Water chemistry, water isotopic composition and carbon isotopic composition of DOC

Samples were clearly differentiated by water chemistry, revealing two major endmembers related to fine- and coarse-textured soils ("Plateau" vs. "Upland"; Figure 2, Supporting Information Table S1), and one intermediate site ("Valley"). Acidity (pH), electrical conductivity (EC) and concentrations of dissolved organic carbon (DOC) were 3.7 ± 0.1 , $49 \pm 6 \mu\text{S cm}^{-1}$ and $37.7 \pm 7.54 \text{ mg C l}^{-1}$ in Campina samples (Figure 2a – c), and some samples at Reserva Cuieiras

346 indicated the same tendency (PR7, PR8, PR9), suggesting common WSE water properties.
347 Samples from the plateau environment showed slightly higher pH, and low EC and DOC levels
348 (4.5 ± 0.2 , $12 \pm 2 \mu\text{S cm}^{-1}$ and $0.9 \pm 0.6 \text{ mg C l}^{-1}$). Three samples from Reserva Cuieiras (PR10,
349 PT06, and stream RA, shown by open symbols) indicated “intermediate” levels of all three
350 parameters (pH 3.9 – 4.3, EC 14 – 29 $\mu\text{S cm}^{-1}$ and 2.3 – 31.3 mg C l^{-1}). Magnitudes and
351 correlations of acidity (pH), electrical conductivity (EC), and dissolved organic carbon (DOC)
352 concentration agree with previous reports from similar environments, showing a strong positive
353 correlation of EC, proton concentration, and DOC (Do Nascimento et al., 2008; Bardy et al.,
354 2011; Monteiro et al., 2014). Plateau samples showed higher pH values and stronger variability,
355 and the absence of acidic DOC (Do Nascimento et al., 2004; Do Nascimento et al., 2008). DOC
356 concentrations of Rio Açu (RA) were low compared to surrounding valley piezometers. In
357 contrast, DOC concentration of the stream (“Rio Campina”, RC) draining the upland site
358 revealed similarly high DOC levels as the surrounding piezometers. Despite similarities in water
359 chemistry of WSE samples (non-intermediate “Valley” and “Upland” samples, Figure 2a-c),
360 differences were most apparent in water isotopic composition, and radiocarbon content of DOC
361 (Figure 2d-e, Supporting Information Figure S2, Supporting Information Figure S3). All samples
362 plotted on the local meteoric water line, resembling the range of expected isotopic composition
363 found in the region. However, water was overall lighter as compared to local precipitation in
364 previous years (Supporting Information Figure S2). Despite regional climatological effects, sites
365 differed significantly in water composition. Campina samples showed a very homogenous water
366 isotopic composition except for the most upslope sample P2. In comparison, samples from
367 Reserva Cuieiras showed consistently lighter (more negative) $\delta^2\text{H}$ and $\delta^{18}\text{O}$ values; plateau
368 samples were lightest and showed a very homogenous water isotopic composition. The valley

samples indicated variation in isotopic composition within the range of the other samples. While the $\delta^{13}\text{C}$ composition of all DOC extracts was roughly similar (average $\pm$ standard deviation: -29.40 ± 0.45 ‰, $n=13$) and reflected a typical C_3 source, their calibrated radiocarbon ages were young (less than 2 to 13 years old at maximum) and differed markedly by five years between both sites, with the upland site showing older DOC (Figure 2e).

3.2 DOM characterization by ultrahigh-resolution mass spectrometry

The analysis of molecular DOM composition reflected the separation of samples based on water and DOC properties presented in the previous section (Figure 3a). The PCoA separated samples into two main clusters (“plateau” and “WSE”) on the first coordinate, which held 78% of molecular variation. Plateau-derived DOM was thus most dissimilar from WSE-DOM, in line with a major control of DOC properties and abundance by soil texture (Remington et al., 2007). Consequently, PCoA 1 was linked to significant (Pearson’ r , $p < 0.05$) trends in pH, EC, and DOC (not shown). In line with more subtle differences between WSE samples, the overall explained variability of the second coordinate was smaller (10%). However, WSE sites were clearly separated, and molecular trends thus paralleled differences in water isotopes and radiocarbon age. Moreover, the separation of WSE-DOM by sites suggested differing trajectories of DOM processing that seemed to converge to a common DOM composition (note the close proximity of samples P7 and PR7 in Figure 3a). Several significant trends in molecular indices derived from the DOM data paralleled the PCoA separation (Figure 3b-d). The used descriptors – a-priori defined molecular groups and formula classes that aggregate molecular composition information – are described in Supporting Information Table S2, and data is available online from <https://doi.org/10.1594/PANGAEA.922606> (Supporting Information Data Set S1). WSE-DOM was more oxidized (higher O/C and NOSC), less saturated (lower H/C, higher DBE) and

more aromatic (higher AImod) than plateau DOM (Figure 3b). Consequently, a-priori-defined molecular groups reflected those trends (Figure 3c): Aromatic and oxidized groups (polyphenols, black carbon-like, carbohydrate-like, and O-rich highly unsaturated formulae) were more abundant in WSE samples, while aliphatic, less oxidized groups (Oxygen-poor highly unsaturated markers, unsaturated aliphatics, and peptide-like compounds) were dominant in plateau DOM. Average numbers of C, O, and N atoms per formula followed these major trend as well (Figure 3b, d). Simple oxidized formulae (CHO) dominated the upland WSE cluster while the percentage of N- and S- containing formulae (CHNO, CHOS, CHNOS) were more abundant in plateau and plateau-influenced (“intermediate”) valley WSE samples. The percentage of CHNO formulae also differentiated WSE-DOM from upland and valley sites on PCoA2. Besides the effect of N-containing formulae, upland WSE-DOM was also heavier in terms of molecular weight (MW, Figure 3b) due to more C and O atoms per molecular formula. These findings are in line with known bulk characteristics of DOC endmembers from soils, groundwater, and rivers in the region (Leenheer, 1980; McClain et al., 1997; Remington et al., 2007). In a next step we thus extracted the molecular markers that caused the separation of DOM from different plateau and whitesand environments.

As expected from the results of the gradient analysis, we found clear sets of markers for each DOM type that paralleled trends of weight-averaged molecular indices (Figure 3b-d). To this end, we separated unique from shared (“common”) markers, and thus excluded non-informative formulae (Figure 4, Supporting Information Figure S4, and Supporting Information Figure S5). It is important to note that our definition of “unique” and “common” relates to significant differences in abundance. In fact, most molecular formulae were shared based only on presence (30% of all formulae shared among all three ecosystems, and 83% among WSEs; Supporting

Information Figure S6). However, besides subsets of formulae indicating less overlap (CHNO, CHOS: Supporting Information Figure S6; BC, PP, HU: Supporting Information Figure S7), ecosystems differed significantly in abundances of formulae, giving rise to individual “molecular fingerprints” (Supporting Information Figure S5). Markers of plateau DOM showed a narrow mass and chemical space distribution as assessed by van Krevelen diagrams (formulae centered at m/z 425, $O/C < 0.5$, and $H/C > 1$). They were also characterized by high numbers of N- and S-containing formulae (63% of all markers) as compared to “simpler” CHO formulae (Figure 4a-c, and Supporting Information Figure S5a). Unsaturated, relatively less oxidized nitrogen-containing formulae classified as “highly unsaturated” compounds (“HU”, i.e., lignin-like formulae, or carboxyl-rich alicyclic molecules, CRAM) were the most dominant group of markers, representing ~50% of all plateau markers. In contrast, molecular markers common to both WSE-DOM types were dominated by CHO formulae and showed a broad chemical space distribution with molecular weights up to m/z 800, and a distinct center at low m/z (~ 200 Da; Supporting Information Figure S4a-c). The most distinct WSE-DOM markers were oxidized, phenolic/ aromatic CHO formulae (74% belonging to molecular groups “orHU”, “PP”, or “BC”; and 64% belonging to “CHO” class).

We observed clear molecular differences between WSE-DOM from the Cuieiras valley and upland Reserva Campina, as suggested by the explorative PCoA (PCoA 2 in Figure 3a), and unique markers of each WSE reflected this divergence (Figure 4d-f, and g-i). This differentiation was most apparent through a sharp “cutoff” at ~ m/z 400 (Figure 4d, g). Reserva Cuieiras valley samples were characterized by lower-molecular-weight N-containing formulae which represented 62% of all valley markers (weighted average mass ~ m/z 300; Figure 4d, f, and Supporting Information Figure S5b). Despite their low mass, these formulae showed wide

distribution in chemical space, mostly belonging to the classes of highly unsaturated and polyphenolic compounds (“HU”, 47% of all markers, and “PP”, 20%; Figure 4e). Upland WSE-DOM, on the other hand, was differentiated by a confined cluster of CHO formulae with higher average mass ($\sim m/z$ 425). Indicative upland CHO formulae concentrated in a chemical space defined by $O/C > 0.3$ and $H/C < 1$ and showed distribution across a wide mass range, high aromaticity, and high oxidation state (Figure 4h, i). All in all, 57% of upland WSE markers belonged to the molecular groups of “orHU”, “PP”, or “BC”, and 83% were classified “CHO” formulae.

3.3 Molecular links between terrestrial ecosystems and the aquatic continuum

We compared the sets of distinct DOM markers of the valley and upland WSEs to known Rio Negro markers in order to gain qualitative insight into the contribution of headwater aquifer and stream DOM to the “integrated” watershed signal of the lower Rio Negro (Figure 5). Although many CHNO and CHOS formulae were part of the specific sets of signals in both WSEs, the overlap to Rio Negro markers was restricted to CHO formulae. The two WSEs differed in terms of potential molecular links. The cluster of heavy, highly oxidized, and aromatic upland WSE markers indicated consistent overlap with known Rio Negro markers (Figure 5). Overlapping sets of formulae appeared in a confined area of the Van Krevelen plot (boxes in Figure 5b, d; ranges: H/C 0.4 – 0.8, O/C 0.4 – 0.8), suggesting robust matching. However, the overlap was caused by different sets of Rio Negro markers, as indicated by the formula’s m/z (Figure 5c, f). One study (Gonsior et al., 2016) revealed molecular links with markers of higher mass (m/z 350 – 650, Figure 5c), while another (Simon et al., 2019) found them in the lower mass range (m/z 200 – 500, Figure 5f). Nevertheless, all these markers were specific to the upland WSE and

restricted to a common type of chemistry, and valley WSE markers contributed to a much lower degree.

4 Discussion

4.1 Site characterization: Biogeochemistry of water and carbon

The general analysis of water chemistry showed clear differences among samples, revealing two major endmembers and one intermediate environment (Figure 2). Strongest differences were apparent between fine- and coarse-textured soils (“Plateau” vs. “Upland”; Figure 2, Supporting Information Table S1). Valley samples PR10, PT06, and stream RA indicated “intermediate” levels in terms of pH, electrical conductivity and DOC concentration and were likely plateau-influenced, i.e., affected by mixing due to plateau water supply, or dilution from precipitation in case of stream RA at the time point of sampling. In contrast, similarly high DOC concentrations of all samples from Reserva Campina draining the upland site indicated direct contact between stream and surrounding soils at the time point of sampling (McClain et al., 1997; Zanchi et al., 2015). Local differences in ecohydrology can explain the heavier isotopic composition of groundwater and stream water in the drier upland Reserva Campina (Supporting Information Figure S2; Zanchi et al., 2014; Zanchi et al., 2015), and the lighter isotopic composition of water in the valley and its adjacent plateaus (Leopoldo et al., 1982; Kunert et al., 2017). While isotopically heavy samples like upland P2 (most upslope) and valley PR8/ PR9 (high water levels) likely reflected the imprint of recent precipitation, groundwater replenished during previous wet seasons can explain the lighter water isotopic composition (Leopoldo et al., 1982; Tomasella et al., 2007; Zhang et al., 2009; Miguez-Macho and Fan, 2012; Jasechko and Taylor, 2015; Zanchi et al., 2015; Supporting Information Figure S3). Radiocarbon content differed among sites but indicated very young DOC age. Young radiocarbon ages of tropical DOC are

frequently reported from aquatic (Mayorga et al., 2005; Moyer et al., 2013; Ward et al., 2013) and soil systems (James et al., 2019) and are explained by strong linkages between C fixation, DOC release, and nutrient recycling (Mayorga et al., 2005). The slight but consistent differences in DOC radiocarbon ages between the drier upland and the wetter valley system likely relate to processes that govern the short-term (years to decades) turnover of organic matter, for example litter turnover and soil respiration rates (Zanchi et al., 2011; Zanchi et al., 2014; Zanchi et al., 2015).

4.2 *DOM properties and fingerprints reflect biogeochemical differences among ecosystems*

Molecular analysis of DOM strongly reflected both the major distinction between ecosystem types by water chemistry (plateau vs. WSE-type ecosystems) and the minor distinction between WSE sites by water isotopic composition and ^{14}C age (Figure 3). This allowed us to extract sets of unique markers for each environment (Figure 4, Supporting Information Figure S4). The distinct markers of plateau and WSE settings documented a major texture effect on DOM properties: Fine-textured soils can cause longer water retention and contact times between minerals, microbes, and water (Marques et al., 2004; Remington et al., 2007) and may thus favor overall lower DOC levels due to intensified decomposition (Marques et al., 2010), explaining low DOC concentrations. Clay particles are also often associated with N-containing compounds (Chassé et al., 2015; Newcomb et al., 2017), and newly synthesized, larger and N-containing microbial compounds can also become dominant during decomposition (Roth et al., 2019), thus leading to higher CHNO compound abundances and diversity. Previous reports on narrow C/N ratios of DOC (~10) in plateau soils support these general findings (McClain et al., 1997). In turn, higher DOC concentration and the dominance of mainly CHO type, aromatic, and oxidized WSE markers may thus represent initial stages of decomposition that also agree with reports on

wider C/N ratios in WSE sites (>15 , up to 60; McClain et al., 1997). These results demonstrate the overall importance of WSEs for the amount and quality of exported terrestrial DOC in the lower Rio Negro basin.

Distinct differences in WSE-DOM composition between sites may reflect ecohydrological differences linked to the frequency of drying/ rewetting events (non-saturated/ saturated conditions), which was also suggested by water isotopic composition and radiocarbon data. We expected that valley WSE-DOM fingerprints would reflect in part the lateral flows from adjacent plateau environments, and the presence of large amounts of indicative N-containing markers with relatively high saturation, low oxidation and low molecular weight supports this assumption (compare Figure 4b,c and e,f). Large differences in DOM composition, however, indicate a loss of the plateau-derived DOM signature upon transit through the valley WSE system before groundwater enters the stream. This finding is remarkable because the riparian zone concentrates the water flux that is sourced from the surrounding plateaus (Miguez-Macho and Fan, 2012a). Simultaneously, this water flux drives the continuous export of young dissolved organic matter from the riparian valleys (Ledesma et al., 2015). However, permanent saturation, i.e., stagnating waters, seem to contribute to the preservation of a wide suite of organic compounds as opposed to the well-drained upland Campina site, and could thus explain the presence of unique, and mainly N-containing, valley WSE-DOM markers. Oxygen depletion in stagnating groundwater limits the turnover of reduced organic matter (Boye et al., 2017). The prevalence of low molecular weight ($m/z < 450$) N- and S-containing molecular formulae with a relatively high degree of saturation ($H/C > 1$) and a low degree of oxygenation ($O/C < 0.6$) could thus also be due to limited DOM uptake caused by unfavorable environmental conditions (low oxygen, low pH, high concentrations of phenolics; Bardy et al., 2011). In contrast, upland WSE-DOM

contained unique, highly oxidized, aromatic, and phenolic CHO formulae (Figure 4g-i). These molecules likely represent the initial decomposition products of plant material containing large portions of lignin, cellulose, tannin, flavonoids, and terpenoids. Under non-stagnating, well-drained conditions, sandy soils favor the escape of such surface signals to streams because of the low sorption capacity of soil (Remington et al., 2007). However, fast export contrasts with older calibrated radiocarbon ages of DOC at the upland WSE (~11 years) as compared to the valley WSE site. This delay can be explained by slower litter turnover and reduced CO₂ efflux rates as discussed above (Zanchi et al., 2011; Zanchi et al., 2014). Highly acidic conditions, periods of drought, and an adapted plant community emerge as main drivers of such ecosystem-level differences. It is of special interest that fungi are known to remain active under dry and acidic conditions (Rousk et al., 2010; Vasco-Palacios et al., 2018). Fungal enzymes can potentially alter DOM towards higher-molecular weight, aromatic, and oxidized structures (Waggoner et al., 2015; Zavarzina et al., 2018), all of which seem to be occurring uniquely in upland WSE soils.

4.3 Upland Campina WSEs are a potential source of indicative Rio Negro markers

We found a clear distinction between potential sources of known sets of Rio Negro markers that could be suitable for further efforts to quantify terrestrial transfer of organic carbon to the regional aquatic continuum and beyond (Figure 5). While the matching with plateau markers was negligible, the two WSE environments clearly differed in terms of matching markers. Independent of the two sets of known Rio Negro DOM markers, a cluster of relatively heavy, highly oxidized, and aromatic upland WSE markers indicated matching (“overlap”) with known Rio Negro markers (Figure 5b, e). Differences in overlap among the both available Rio Negro datasets (Gonsior et al., 2016; Simon et al., 2019) does not diminish the consistent match in terms of chemical space (black boxes in Figure 5b, e). Fine-tuning of measurement conditions

can easily affect the instrument response in terms of ion abundance patterns and thus, m/z range (Hawkes et al., 2016; Simon et al., 2018; Hawkes et al., 2020). Despite sample set and lab/instrument effects, measurement settings were relatively similar among the two available studies and ours, thus supporting the general overlap of the upland WSE marker and Rio Negro marker sets in chemical space (oxidation and saturation state; Supporting Information Text S1). Accumulation time of ions before FT-MS analysis is one factor that explains m/z shifts (Hawkes et al., 2016; Simon et al., 2018). However, sampling criteria such as location, season, scale, and coverage will also have a strong influence on the resulting sets of markers. While it is not surprising to see large differences in marker sets (Supporting Information Figure S8 and Supporting Information Figure S9), it is remarkable to find small but consistent overlap between studies, pointing toward subsets of markers with specific oxidation and saturation states. It is known that different instruments capture gradients of variation across a sample set even at very slight compositional differences (Hawkes et al., 2016), as may be expected when samples originate from similar environmental contexts as in our study. It is thus highly encouraging to find consistent and robust molecular overlap in three independent FT-MS datasets, being in line with long-standing hypotheses of landscape functioning in the Rio Negro basin (Leenheer, 1980; Goulding et al., 1988), namely that whitesand ecosystems are potential sources of indicative Rio Negro markers. Counter-intuitively, the highly indicative set of upland WSE markers were not only found to overlap with Rio Negro samples from the proximity of its draining higher-order river, the Rio Tarumã Açu (Simon et al., 2019) but also in samples upstream (Gonsior et al., 2016; compare Figure 1). This observation implies that similar high-molecular-weight oxidized aromatic compounds are exported upstream of Novo Airão, possibly by other upland, Campina-

covered WSE systems. The DOM fingerprints of these other (extensive) WSE systems, especially in the upper Rio Negro basin (Adeney et al., 2016), remain to be revealed.

To identify potentially robust candidate markers, we queried the nine upland whitesand Rio Negro markers that consistently matched with the two independent lists of Rio Negro markers (“Upland overlap”, red squares in Figure 5) in PubChem to obtain potential structure suggestions (Supporting Information Table S3). A query with the ChEBI (chemical entities of biological interest) database yielded only two suggestions that were also found in PubChem. The structural suggestions thus represent only a first indication of potential structures and not necessarily structures that were identified in DOM or organisms. Pubchem yielded up to 108 hits for formula $C_{10}H_6O_6$ and five at minimum for formula $C_{11}H_6O_8$ (numbers after exclusion of twelve ions, one hydrate and six stereoisomers from the list). In general, structures were highly oxidized, showing on average 2.5 carbonyl groups, 0.97 ether bonds, and 0.65 lactone groups per molecule across formulae. All formulae were classified as Black-Carbon like or polyphenol-like (BC, PP). On average, 68% of suggested structures per formula featured at least one heterocyclic ring, with five and six-membered rings dominating (each 30% on average). Functional group count was strongly correlated with mass ($R^2 = 0.84$, range = 5.2 – 9.7), same as numbers of carbon double bonds (0.86, 3.2 – 8.4) aromatic rings per molecule (0.87, 0.6 – 2.6) and thus, DBE (0.85, 8 – 16). Suggestions of smaller molecules (< 270 Da, with less than 15 C atoms) were by tendency dominated by naphthalene and chromene structures that are oxygen-poor scaffolds consisting of two rings. Larger molecules > 320 Da (with >15 C atoms) in contrast showed more hits involving at least one benzofuran, phenol or benzenecarboxylic scaffold, which all show a higher degree of oxidation and functionalization. Chromenes and benzofurans are well known photo- and bioactive plant metabolites (Towers and Hudson, 1987). Knowledge of these trends and

tendencies in molecular properties may help to develop tailored chromatographic techniques (LC-MS) for the targeted analysis of these biological relevant ecosystem markers in future (Petras et al., 2017). Structural data used for the above analyses is openly available from <https://doi.org/10.1594/PANGAEA.922606> (Supporting Information Data Set S1, Supporting Information Data Set S3).

Groundwater DOM markers revealed a potential direct link between ecosystem-specific headwater and ecosystem-integrated downstream signals in the Rio Negro basin. The consistent overlap of upland WSE and Rio Negro markers was unexpected because our initial hypothesis, namely that typical valley WSE's will show stronger, or at least similar, matching with Rio Negro markers, had to be refused. Highly specific, tannin-like aromatic DOM compounds of the smaller upland WSE watershed represent a potential molecular link between headwaters and the Rio Negro. This finding opposes the assumed role of riparian valley WSEs as main sources of DOM in the Rio Negro watershed (McClain et al., 1997; Bardy et al., 2011). Although Bardy and coworkers showed that upland Campina forests areas produce a marked DOM signal traceable to rivers, others found that riparian and wetland WSEs along rivers contributed mainly to DOC exports of Amazonian and other watersheds (Dosskey and Bertsch, 1994; McClain et al., 1997; Remington et al., 2007). Our results suggest that there is more detail to add to this simple model. Small watersheds with strong terrestrial-lotic linkages may leave a more significant downstream imprint as expected from absolute annual discharge, areal extent, or DOC export. Our results show that highly specific forest ecosystems potentially leave a distinct imprint within the Rio Negro's exported DOM, although contributing only secondarily to the overall fluxes of water, and maybe, DOC. The smaller upland WSE watershed at Reserva Campina shows a two times lower average annual stream discharge (based on the catchment area) but a two-three times

higher annual DOC export (Monteiro et al., 2014; Zanchi et al., 2015). This discrepancy demonstrates that water and carbon cycles may be decoupled on the molecular level and the watershed scale, despite the generally accepted transport-limited, or chemostatic, DOC behavior (Musolff et al., 2017; Zarnetske et al., 2018). Ecosystems may show different behavior in terms of discharge and DOC export (Webb et al., 2018), or release of ecosystem markers (Wagner et al., 2019). Taking into account that hydromorphic soils in this region are thought to cover more than 40% of area (Junk, 1993), their missing imprint is evidence for the importance of unique ecosystems in control of indicative DOM characteristics within the land-to-ocean continuum. Riparian valley corridors likely contribute the major part of the annual DOC export in this region due to constant water supply from adjacent plateaus and their wider spatial distribution (Remington et al., 2007; Miguez-Macho and Fan, 2012a; Miguez-Macho and Fan, 2012b), but do only contribute general DOM markers that likely reflect processes operating in many watersheds. In conclusion, our results indicate that riparian valleys may contribute less to the specific signal that discerns the Rio Negro from other watersheds on the molecular level. These specific markers show the potential for qualitative ecosystem recognition further downstream, and may thus serve as new proxies of land-derived primary production in the Rio Negro basin if properly calibrated.

5 Conclusion

This study investigated potential molecular links between indicative DOM markers at two main stages of the land-to-ocean-continuum, headwater catchments and river basins. New sets of markers are needed to better constrain variations in land-derived DOM exports and losses within aquatic systems, especially in the tropics, which account for 62% of riverine DOC exports. We provide molecular evidence of long-assumed hypothetical links drawn from the simultaneous

occurrence of tropical whitesand ecosystems (WSEs) and blackwater rivers at the example of the Rio Negro basin (Goulding et al., 1988, and references therein). Unique whitesand ecosystems contribute mainly to the fast export of DOC, while it is retained and decomposed in the highly productive, widespread plateau systems. Regardless of the wide occurrence of whitesand ecosystems and the known variation in WSE characteristics across the Amazon, there is little appreciation of this knowledge in models of DOC export and information content. This gap is due to a lack of qualitative DOC data. By using FT-MS techniques, we provide unique markers of whitesand ecosystems that can serve as future constraints on the terrestrial portion of aquatic DOC export. However, for such applications further calibration studies are necessary.

Against initial expectation, large valley WSEs may contribute only little to the flux of soluble ecosystem markers found at the river basin scale. In contrast, highly specific upland Campina WSEs emerged as their potential source. The twofold larger annual discharge and hundred-fold larger size of the Cuieiras watershed seemingly do not result in the transfer of DOM markers downstream, as expected from conservative (chemostatic) DOC leaching of dominant source layers. Instead, estimates suggest that the two-three times higher annual DOC export of the smaller upland watershed leaves a stronger imprint in terms of traceable WSE markers. This discrepancy stresses the importance of specific ecosystems for DOM information content at higher-order stages of the aquatic continuum, and how it may inform hydrological models that include qualitative DOC data. Future research needs to assess the influence of the large bandwidth of WSE types across the Rio Negro basin, and relate their environmental characteristics (climate, soil, vegetation) and geographical extent with information on quantity (DOC) and quality (molecular composition) of the exported DOM.

The molecular composition of DOM at different stages of the aquatic continuum emerges as a qualitative measure of DOM exports that complements quantitative DOC data. Our study shows the gained information that can be drawn from comparative FT-MS studies when complemented with other types of ecosystem information, especially information on the isotopic composition of water and organic carbon. Integration and nesting of DOM data show great prospects to bridge traditional gaps between soil science, limnology, and hydrology. The interconnected nature and multivariate complexity of DOM shows high potential to allow studying multiple source contributions and processes simultaneously. However, improvements in data integration across FT-MS platforms are required to reveal this information properly, and to reach robust conclusions about sources, fate, and identity of new sets of ecosystem markers in DOM. For this it will also be central to calibrate novel ecosystem markers with environmental variables such as DOC export, discharge, or ecosystem productivity.

Acknowledgments

We thank Axel Steinhof and Heike Machts for radiocarbon analyses and data, and Heiko Moossen for support in analyses of water isotopic composition and feedback on earlier versions of the manuscript. This work was accomplished in the framework of the Amazon Tall Tower Observatory (ATTO). We acknowledge funding and support from the Max-Planck-Gesellschaft (MPG), German Bundesministerium für Bildung und Forschung (BMBF), Brazilian Ministry of Science, Technology, Innovation and Communications (MCTIC), Amazonas State Foundation for Research (FAPEAM), Large-scale Biosphere-Atmosphere Experiment of Brazil's National Institute for Amazon Research (LBA/ INPA), Uatumã Sustainable Development Reserve of Amazonas State's Secretariat of Sustainable Development (SDS/ CEUC/ RDS-Uatumã), and São Paulo Research Foundation (FAPESP). CS received a Ph.D. stipend from the International Max

Planck Research School for Global Biogeochemical Cycles (IMPRS-gBGC). We are also grateful to Deutsche Forschungsgemeinschaft (DFG) for financial support as part of CRC 1076 “AquaDiva”.

Declaration of potential conflicts of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data availability

All data necessary to reproduce the findings reported in this manuscript are available online free of charge via <https://doi.org/10.1594/PANGAEA.922606>.

Supplementary Material

29 pages, containing 9 Tables, 17 Figures and 17 References.

References

- Abbott B. W., Gruau G., Zarnetske J. P., Moatar F., Barbe L., Thomas Z., Fovet O., Kolbe T., Gu S., Pierson-Wickmann A. C., Davy P. and Pinay G. (2018) Unexpected spatial stability of water chemistry in headwater stream networks. *Ecol. Lett.* **21**, 296–308.
- Adeney J. M., Christensen N. L., Vicentini A. and Cohn-Haft M. (2016) White-sand ecosystems in Amazonia. *Biotropica* **48**, 7–23.
- Alvarez-Cobelas M., Angeler D. G., Sánchez-Carrillo S. and Almendros G. (2012) A worldwide view of organic carbon export from catchments. *Biogeochemistry* **107**, 275–293.
- Bardy M., Derenne S., Allard T., Benedetti M. F. and Fritsch E. (2011) Podzolisation and exportation of organic matter in black waters of the Rio Negro (upper Amazon basin, Brazil). *Biogeochemistry* **106**, 71–88.
- Benk S. A., Li Y., Roth V.-N. and Gleixner G. (2018) Lignin Dimers as Potential Markers for

711 ¹⁴C-young Terrestrial Dissolved Organic Matter in the Critical Zone. *Front. Earth Sci.*, 1–
 712 9.
 713 Bernhardt E. S., Heffernan J. B., Grimm N. B., Stanley E. H., Harvey J. W., Arroita M., Appling
 714 A. P., Cohen M. J., McDowell W. H., Hall R. O., Read J. S., Roberts B. J., Stets E. G. and
 715 Yackulic C. B. (2018) The metabolic regimes of flowing waters. *Limnol. Oceanogr.* **63**, 99–
 716 118.
 717 Bowen G. J. and Revenaugh J. (2003) Interpolating the isotopic composition of modern meteoric
 718 precipitation. *Water Resour. Res.* **39**, 1–13.
 719 Boye K., Noël V., Tfaily M. M., Bone S. E., Williams K. H., Bargar J. R. and Fendorf S. (2017)
 720 Thermodynamically controlled preservation of organic carbon in floodplains. *Nat. Geosci.*
 721 **10**, 415–419.
 722 Cao X., Aiken G. R., Butler K. D., Huntington T. G., Balch W. M., Mao J. and Schmidt-Rohr K.
 723 (2018) Evidence for major input of riverine organic matter into the ocean. *Org. Geochem.*
 724 **116**, 62–76.
 725 Chassé A. W., Ohno T., Higgins S. R., Amirbahman A., Yildirim N. and Parr T. B. (2015)
 726 Chemical force spectroscopy evidence supporting the layer-by-layer model of organic
 727 matter binding to iron (oxy)hydroxide mineral surfaces. *Environ. Sci. Technol.* **49**, 9733–
 728 9741.
 729 Coynel A., Seyler P., Etcheber H., Meybeck M. and Orange D. (2005) Spatial and seasonal
 730 dynamics of total suspended sediment and organic carbon species in the Congo River.
 731 *Global Biogeochem. Cycles* **19**, 1–17.
 732 Creed I. F., McKnight D. M., Pellerin B. A., Green M. B., Bergamaschi B. A., Aiken G. R.,
 733 Burns D. A., Findlay S. E. G., Shanley J. B., Striegl R. G., Aulenbach B. T., Clow D. W.,

734 Laudon H., McGlynn B. L., McGuire K. J., Smith R. A. and Stackpoole S. M. (2015) The
 735 river as a chemostat: fresh perspectives on dissolved organic matter flowing down the river
 736 continuum. *Can. J. Fish. Aquat. Sci.* **72**, 1272–1285.

737 Dai M., Yin Z., Meng F., Liu Q. and Cai W.-J. (2012) Spatial distribution of riverine DOC inputs
 738 to the ocean: An updated global synthesis. *Curr. Opin. Environ. Sustain.* **4**, 170–178.

739 Demarchi L. O., Scudeller V. V., Moura L. C., Dias-Terceiro R. G., LOPES A., Wittmann F. K.
 740 and Piedade M. T. F. (2018) Floristic composition, structure and soil-vegetation relations in
 741 three white-sand soil patches in central Amazonia. *Acta Amaz.* **48**, 46–56.

742 Dittmar T., Koch B., Hertkorn N. and Kattner G. (2008) A simple and efficient method for the
 743 solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnol.*
 744 *Oceanogr. Methods* **6**, 230–235.

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

747 Drake T. W., Guillemette F., Hemingway J. D., Chanton J. P., Podgorski D. C., Zimov N. S. and
 748 Spencer R. G. M. (2018a) The ephemeral signature of permafrost Carbon in an Arctic
 749 fluvial network. *J. Geophys. Res. Biogeosciences* **123**, 1475–1485.

750 Drake T. W., Van Oost K., Barthel M., Bauters M., Hoyt A. M., Podgorski D. C., Six J., Boeckx
 751 P., Trumbore S. E., Cizungu Ntaboba L. and Spencer R. G. M. (2019) Mobilization of aged
 752 and biolabile soil carbon by tropical deforestation. *Nat. Geosci.* **12**, 541–546.

753 Drake T. W., Raymond P. A. and Spencer R. G. M. (2018b) Terrestrial carbon inputs to inland
 754 waters: A current synthesis of estimates and uncertainty. *Limnol. Oceanogr. Lett.* **3**, 132–
 755 142.

756 Gehre M., Geilmann H., Richter J., Werner R. A. and 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.

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

Goulding M., Carvalho M. L. and Ferreira E. G. (1988) Blackwaters. In *Rio Negro, rich life in poor water. Amazonian diversity and foodchain ecology as seen through fish communities* (eds. M. Goulding, M. L. Carvalho, and E. G. Ferreira). SPB Academic Publishing, The Hague. pp. 29–36.

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.

Guinoiseau D., Bouchez J., Gélabert A., Louvat P., Filizola N. and Benedetti M. F. (2016) The geochemical filter of large river confluences. *Chem. Geol.* **441**, 191–203.

Hawkes J. A., D’Andrilli J., Agar J. N., Barrow M. P., Berg S. M., Catalán N., Chen H., Chu R. K., Cole R. B., Dittmar T., Gavard R., Gleixner G., Hatcher P. G., He C., Hess N. J., Hutchins R. H. S., Ijaz A., Jones H. E., Kew W., Khaksari M., Lozano D. C. P., Lv J., Mazzoleni L., Noriega-Ortega B., Osterholz H., Radoman N., Remucal C. K., Schmitt N. D., Schum S., Shi Q., Simon C., Singer G., Sleighter R. S., Stubbins A., Thomas M. J., Tolic N., Zhang S., Zito P. and Podgorski D. C. (2020) An international laboratory comparison of dissolved organic matter composition by high resolution mass spectrometry: Are we getting the same answer? *Limnol. Oceanogr. Methods* **accepted**, doi: 10.1002/lom3.10364.

Hawkes J. A., Dittmar T., Patriarca C., Tranvik L. and Bergquist J. (2016) Evaluation of the

Orbitrap Mass Spectrometer for the Molecular Fingerprinting Analysis of Natural Dissolved Organic Matter. *Anal. Chem.* **88**, 7698–7704.

Hertkorn N., Harir M., Koch B. P., Michalke B. and 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.

Herzprung P., Hertkorn N., von Tümpling W., Harir M., Frieze K. and 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.

Huang T. H., Fu Y. H., Pan P. Y. and Chen C. T. A. (2012) Fluvial carbon fluxes in tropical rivers. *Curr. Opin. Environ. Sustain.* **4**, 162–169.

Hutchins R. H. S., Aukes P., Schiff S. L., Dittmar T., Prairie Y. T. and 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.

James J. N., Gross C. D., Dwivedi P., Myers T., Santos F., Bernardi R., Fidalgo de Faria M., Amaral Guerrini I., Harrison R. and 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.

Janzen D. H. (1974) Tropical blackwater rivers, animals, and mast fruiting by the Dipterocarpaceae. *Biotropica* **6**, 69–103.

Jasechko S. and Taylor R. G. (2015) Intensive rainfall recharges tropical groundwaters. *Environ. Res. Lett.* **10**, 124015.

Jehn F. U., Bestian K., Breuer L., Kraft P. and Houska T. (2020) Using hydrological and climatic

803 catchment clusters to explore drivers of catchment behavior. *Hydrol. Earth Syst. Sci.* **24**,
804 1081–1100.

805 Junk W. J. (1993) Wetlands of tropical South America. In *Wetlands of the world: Inventory,*
806 *ecology and management.* (eds. D. F. Whigham, D. Dykyjová, and S. Hejný). Springer
807 Science+Business Media B.V., Heidelberg. pp. 679–739.

808 Junk W. J., Piedade M. T. F., Schöngart J., Cohn-Haft M., Adeney J. M. and Wittmann F. (2011)
809 A classification of major naturally-occurring Amazonian Lowland Wetlands. *Wetlands* **31**,
810 623–640.

811 Kellerman A. M., Guillemette F., Podgorski D. C., Aiken G. R., Butler K. D. and Spencer R. G.
812 M. (2018) Unifying Concepts Linking Dissolved Organic Matter Composition to
813 Persistence in Aquatic Ecosystems. *Environ. Sci. Technol.* **52**, 2538–2548.

814 Klinge H. and Medina E. (1979) Rio Negro caatingas and campinas, Amazonas States of
815 Venezuela and Brazil. In *Ecosystems of the World: Heathlands and related shrublands.*
816 *Ecosystems of the World. A. Descriptive studies* (ed. R. L. Specht). Elsevier, Amsterdam.
817 pp. 483–488.

818 Kunert N., Aparecido L. M. T., Wolff S., Higuchi N., Dos Santos J., De Araujo A. C. and
819 Trumbore S. (2017) A revised hydrological model for the Central Amazon: The importance
820 of emergent canopy trees in the forest water budget. *Agric. For. Meteorol.* **239**, 47–57.

821 Laudon H. and Sponseller R. A. (2018) How landscape organization and scale shape catchment
822 hydrology and biogeochemistry: insights from a long-term catchment study. *Wiley*
823 *Interdiscip. Rev. Water* **5**, e1265.

824 Ledesma J. L. J., Grabs T., Bishop K. H., Schiff S. L. and Köhler S. J. (2015) Potential for long-
825 term transfer of dissolved organic carbon from riparian zones to streams in boreal

826 catchments. *Glob. Chang. Biol.* **21**, 2963–2979.

827 Leenheer J. A. (1980) Origin and nature of humic substances in the waters of the Amazon River
828 Basin. *Acta Amaz.* **10**, 513–526.

829 Leopoldo P. R., Matsui E., Salati E., Franken W. and Ribeiro M. de N. G. (1982) Composição
830 isotópica da água de chuva e da água do solo em floresta amazônica do tipo terra firme,
831 região de Manaus. *Acta Amaz.* **12**, 7–13.

832 Li M., Peng C., Zhou X., Yang Y., Guo Y., Shi G. and Zhu Q. (2019) Modeling global riverine
833 DOC flux dynamics from 1951 to 2015. *J. Adv. Model. Earth Syst.* **11**, 514–530.

834 Lucas Y., Montes C. R., Mounier S., Loustau Cazalet M., Ishida D., Achard R., Garnier C.,
835 Coulomb B. and Melfi A. J. (2012) Biogeochemistry of an Amazonian podzol-ferralsol soil
836 system with white kaolin. *Biogeosciences* **9**, 3705–3720.

837 Luizão R. C. C., Luizão F. J., Paiva R. Q., Monteiro T. F., Sousa L. S. and Kruijt B. (2004)
838 Variation of carbon and nitrogen cycling processes along a topographic gradient in a central
839 Amazonian forest. *Glob. Chang. Biol.* **10**, 592–600.

840 Lv S., Yu Q., Wang F., Wang Y., Yan W. and Li Y. (2019) A synthetic model to quantify
841 dissolved organic carbon transport in the Changjiang River system: Model structure and
842 spatiotemporal patterns. *J. Adv. Model. Earth Syst.* **11**, 3024–3041.

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

846 Marques J. D. de O., Libardi P. L., Teixeira W. G. and Reis A. M. (2004) Estudo de parâmetros
847 físicos, químicos e hídricos de um Latossolo Amarelo, na região Amazônica. *Acta Amaz.*
848 **34**, 145–154.

849 Marques J. D. de O., Luizão F. J., Teixeira W. G., Vitel C. M. and Marques E. M. de A. (2016)
850 Soil organic carbon, carbon stock and their relationships to physical attributes under forest
851 soils in central Amazonia. *Rev. Árvore* **40**, 197–208.

852 Marques J. D. de O., Teixeira W. G., Reis A. M., Cruz Junior O. F., Batista S. M. and Afonso M.
853 A. C. B. (2010) Atributos químicos, físico-hídricos e mineralogia da fração argila em solos
854 do Baixo Amazonas: Serra de Parintins. *Acta Amaz.* **40**, 01–12.

855 Mayorga E., Aufdenkampe A. K., Masiello C. a, Krusche A. V, Hedges J. I., Quay P. D., Richey
856 J. E. and Brown T. A. (2005) Young organic matter as a source of carbon dioxide
857 outgassing from Amazonian rivers. *Nature* **436**, 538–541.

858 McClain M. E., Richey J. E., Brandes J. A. and Pimentel T. P. (1997) Dissolved organic matter
859 and terrestrial-lotic linkages in the central Amazon basin of Brazil. *Global Biogeochem.*
860 *Cycles* **11**, 295–311.

861 McGuire K. J., Torgersen C. E., Likens G. E., Buso D. C., Lowe W. H. and Bailey S. W. (2014)
862 Network analysis reveals multiscale controls on streamwater chemistry. *Proc. Natl. Acad.*
863 *Sci. U. S. A.* **111**, 7030–7035.

864 Medeiros P. M., Seidel M., Niggemann J., Spencer R. G. M., Hernes P. J., Yager P. L., Miller W.
865 L., Dittmar T. and Hansell D. A. (2016) A novel molecular approach for tracing terrigenous
866 dissolved organic matter into the deep ocean. *Global Biogeochem. Cycles* **30**, 1–11.

867 Melack J. M. and Hess L. L. (2010) Remote Sensing of the Distribution and Extent of Wetlands
868 in the Amazon Basin. In *Amazonian Floodplain Forests: Ecophysiology, Biodiversity and*
869 *Sustainable Management. Ecological Studies 210* (eds. W. J. Junk, M. T. F. Piedade, F.
870 Wittmann, J. Schöngart, and P. Parolin). Springer Science+Business Media B.V.,
871 Heidelberg. pp. 43–59.

872 Miguez-Macho G. and Fan Y. (2012a) The role of groundwater in the Amazon water cycle: 1.
873 Influence on seasonal streamflow, flooding and wetlands. *J. Geophys. Res. Atmos.* **117**,
874 D15113.

875 Miguez-Macho G. and Fan Y. (2012b) The role of groundwater in the Amazon water cycle: 2.
876 Influence on seasonal soil moisture and evapotranspiration. *J. Geophys. Res. Atmos.* **117**,
877 D15114.

878 Monteiro M. T. F., Oliveira S. M., Luizão F. J., Cândido L. A., Ishida F. Y. and Tomasella J.
879 (2014) Dissolved organic carbon concentration and its relationship to electrical conductivity
880 in the waters of a stream in a forested Amazonian blackwater catchment. *Plant Ecol.*
881 *Divers.* **7**, 205–213.

882 Montes C. R., Lucas Y., Pereira O. J. R., Achard R., Grimaldi M. and Melfi A. J. (2011) Deep
883 plant-derived carbon storage in Amazonian podzols. *Biogeosciences* **8**, 113–120.

884 Moore S., Evans C. D., Page S. E., Garnett M. H., Jones T. G., Freeman C., Hooijer A., Wiltshire
885 A. J., Limin S. H. and Gauci V. (2013) Deep instability of deforested tropical peatlands
886 revealed by fluvial organic carbon fluxes. *Nature* **493**, 660–663.

887 Moyer R. P., Bauer J. E. and Grottoli A. G. (2013) Carbon isotope biogeochemistry of tropical
888 small mountainous river, estuarine, and coastal systems of Puerto Rico. *Biogeochemistry*
889 **112**, 589–612.

890 Musolff A., Fleckenstein J. H., Rao P. S. C. and Jawitz J. W. (2017) Emergent archetype patterns
891 of coupled hydrologic and biogeochemical responses in catchments. *Geophys. Res. Lett.* **44**,
892 4143–4151.

893 Do Nascimento N. R., Bueno G. T., Fritsch E., Herbillon A. J., Allard T., Melfi A. J., Astolfo R.,
894 Boucher H. and Li Y. (2004) Podzolization as a deferralitization process: A study of an

895 Acrisol-Podzol sequence derived from Palaeozoic sandstones in the northern upper Amazon
896 Basin. *Eur. J. Soil Sci.* **55**, 523–538.

897 Do Nascimento N. R., Fritsch E., Bueno G. T., Bardy M., Grimaldi C. and Melfi A. J. (2008)
898 Podzolization as a deferralitization process: Dynamics and chemistry of ground and surface
899 waters in an Acrisol - Podzol sequence of the upper Amazon Basin. *Eur. J. Soil Sci.* **59**,
900 911–924.

901 Newcomb C. J., Qafoku N. P., Grate J. W., Bailey V. L. and De Yoreo J. J. (2017) Developing a
902 molecular picture of soil organic matter-mineral interactions by quantifying organo-mineral
903 binding. *Nat. Commun.* **8**, 396.

904 Oksanen J. (2010) Multivariate analysis of ecological communities in R: vegan tutorial. , 43.
905 Available at: cc.oulu.fi/~jarioksa/opetus/metodi/vegantutor.pdf [Accessed April 6, 2015].

906 Osterholz H., Singer G., Wemheuer B., Daniel R., Simon M., Niggemann J. and Dittmar T.
907 (2016) Deciphering associations between dissolved organic molecules and bacterial
908 communities in a pelagic marine system. *ISME J.* **10**, 1717–1730.

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

913 Quesada C. A., Lloyd J., Anderson L. O., Fyllas N. M., Schwarz M. and Czimczik C. I. (2011)
914 Soils of Amazonia with particular reference to the RAINFOR sites. *Biogeosciences* **8**,
915 1415–1440.

916 Raeke J., Lechtenfeld O. J., Tittel J., Oosterwoud M. R., Bornmann K. and Reemtsma T. (2017)
917 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.

Raymond P. A. and Spencer R. G. M. (2014) Riverine DOM. In *Biogeochemistry of Marine Dissolved Organic Matter: Second Edition* (eds. D. A. Hansell and C. A. Carlson). Academic Press, Cambridge. pp. 509–533.

Regnier P., Friedlingstein P., Ciais P., Mackenzie F. T., Gruber N., Janssens I. A., Laruelle G. G., Lauerwald R., Luyssaert S., Andersson A. J., Arndt S., Arnosti C., Borges A. V., Dale A. W., Gallego-Sala A., Godd  ris Y., Goossens N., Hartmann J., Heinze C., Ilyina T., Joos F., Larowe D. E., Leifeld J., Meysman F. J. R., Munhoven G., Raymond P. A., Spahni R., Suntharalingam P. and Thullner M. (2013) Anthropogenic perturbation of the carbon fluxes from land to ocean. *Nat. Geosci.* **6**, 597–607.

Remington S. M., Strahm B. D., Neu V., Richey J. E. and 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.

Riedel T. and 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.

Riedel T., Zark M., V  h  t  lo A. V., Niggemann J., Spencer R. G. M., Hernes P. J. and Dittmar T. (2016) Molecular Signatures of Biogeochemical Transformations in Dissolved Organic Matter from Ten World Rivers. *Front. Earth Sci.* **4**, 85.

Roth V.-N., Dittmar T., Gaupp R. and Gleixner G. (2014) Ecosystem-specific composition of dissolved organic matter. *Vadose Zo. J.* **13**.

Roth V.-N., Lange M., Simon C., Hertkorn N., Bucher S., Goodall T., Griffiths R. I., Mellado-V  zquez P. G., Mommer L., Oram N. J., Weigelt A., Dittmar T. and Gleixner G. (2019)

941 Persistence of dissolved organic matter explained by molecular changes during its passage
 942 through soil. *Nat. Geosci.* **12**, 755–761.

943 Rousk J., Bååth E., Brookes P. C., Lauber C. L., Lozupone C., Caporaso J. G., Knight R. and
 944 Fierer N. (2010) Soil bacterial and fungal communities across a pH gradient in an arable
 945 soil. *ISME J.* **4**, 1340–1351.

946 Simon C., Osterholz H., Koschinsky A. and Dittmar T. (2019) Riverine mixing at the molecular
 947 scale – An ultrahigh-resolution mass spectrometry study on dissolved organic matter and
 948 selected metals in the Amazon confluence zone (Manaus, Brazil). *Org. Geochem.* **129**, 45–
 949 62.

950 Simon C., Roth V.-N., Dittmar T. and Gleixner G. (2018) Molecular Signals of Heterogeneous
 951 Terrestrial Environments Identified in Dissolved Organic Matter: A Comparative Analysis
 952 of Orbitrap and Ion Cyclotron Resonance Mass Spectrometers. *Front. Earth Sci.* **6**, 1–16.

953 Sioli H. (1954) Gewässerchemie und Vorgänge in den Böden im Amazonasgebiet.
 954 *Naturwissenschaften* **41**, 456–457.

955 Spencer R. G. M., Kellerman A. M., Podgorski D. C., Macedo M. N., Jankowski K. J., Nunes D.
 956 and Neill C. (2019) Identifying the Molecular Signatures of Agricultural Expansion in
 957 Amazonian Headwater Streams. *J. Geophys. Res. Biogeosciences*, 1637–1650.

958 Steinhof A., Altenburg M. and Machts H. (2017) Sample preparation at the Jena 14C laboratory.
 959 *Radiocarbon* **59**, 815–830.

960 Tiegs S. D., Costello D. M., Isken M. W., Woodward G., McIntyre P. B., Gessner M. O.,
 961 Chauvet E., Griffiths N. A., Flecker A. S., Acuña V., Albariño R., Allen D. C., Alonso C.,
 962 Andino P., Arango C., Aroviita J., Barbosa M. V. M., Barmuta L. A., Baxter C. V., Bell T.
 963 D. C., Bellinger B., Boyero L., Brown L. E., Bruder A., Bruesewitz D. A., Burdon F. J.,

964 Callisto M., Canhoto C., Capps K. A., Castillo M. M., Clapcott J., Colas F., Colón-Gaud C.,
 965 Cornut J., Crespo-Pérez V., Cross W. F., Culp J. M., Danger M., Dangles O., De Eyto E.,
 966 Derry A. M., Villanueva V. D., Douglas M. M., Elosegí A., Encalada A. C., Entekin S.,
 967 Espinosa R., Ethaiya D., Ferreira V., Ferriol C., Flanagan K. M., Fleituch T., Shah J. J. F.,
 968 Barbosa A. F., Friberg N., Frost P. C., Garcia E. A., Lago L. G., Soto P. E. G., Ghate S.,
 969 Giling D. P., Gilmer A., Gonçalves J. F., Gonzales R. K., Graça M. A. S., Grace M.,
 970 Grossart H. P., Guérolde F., Gulis V., Hepp L. U., Higgins S., Hishi T., Huddart J., Hudson
 971 J., Imberger S., Iñiguez-Armijos C., Iwata T., Janetski D. J., Jennings E., Kirkwood A. E.,
 972 Koning A. A., Kosten S., Kuehn K. A., Laudon H., Leavitt P. R., Da Silva A. L. L., Leroux
 973 S. J., LeRoy C. J., Lisi P. J., MacKenzie R., Marcarelli A. M., Masese F. O., McKie B. G.,
 974 Medeiros A. O., Meissner K., Miliša M., Mishra S., Miyake Y., Moerke A., Mombrikotb S.,
 975 Mooney R., Moulton T., Muotka T., Negishi J. N., Neres-Lima V., Nieminen M. L.,
 976 Nimptsch J., Ondruch J., Paavola R., Pardo I., Patrick C. J., Peeters E. T. H. M., Pozo J.,
 977 Pringle C., Prussian A., Quenta E., Quesada A., Reid B., Richardson J. S., Rigosi A.,
 978 Rincón J., Rîşnoveanu G., Robinson C. T., Rodríguez-Gallego L., Royer T. V., Rusak J. A.,
 979 Santamans A. C., Selmeçy G. B., Simiyu G., Skuja A., Smykla J., Sridhar K. R.,
 980 Sponseller R., Stoler A., Swan C. M., Szlag D., Teixeira-De Mello F., Tonkin J. D.,
 981 Uusheimo S., Veach A. M., Vilbaste S., Vought L. B. M., Wang C. P., Webster J. R.,
 982 Wilson P. B., Woelfl S., Xenopoulos M. A., Yates A. G., Yoshimura C., Yule C. M., Zhang
 983 Y. X. and Zwart J. A. (2019) Global patterns and drivers of ecosystem functioning in rivers
 984 and riparian zones. *Sci. Adv.* 5, 1–9.
 985 Tomasella J., Hodnett M. G., Cuartas L. A., Nobre A. D., Waterloo M. J. and Oliveira S. M.
 986 (2007) The water balance of an Amazonian micro-catchment: The effect of interannual

987 variability of rainfall on hydrological behaviour. *Hydrol. Process.* **22**, 2133–2147.

988 Towers G. H. N. and Hudson J. B. (1987) Potentially useful antimicrobial and antiviral
 989 phototoxins from plants. *Photochem. Photobiol.* **46**, 61–66.

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

994 Vasco-Palacios A. M., Hernandez J., Peñuela-Mora M. C., Franco-Molano A. E. and Boekhout
 995 T. (2018) Ectomycorrhizal fungi diversity in a white sand forest in western Amazonia.
 996 *Fungal Ecol.* **31**, 9–18.

997 Waggoner D. C., Chen H., Willoughby A. S. and Hatcher P. G. (2015) Formation of black
 998 carbon-like and alicyclic aliphatic compounds by hydroxyl radical initiated degradation of
 999 lignin. *Org. Geochem.* **82**, 69–76.

1000 Wagner S., Fair J. H., Matt S., Hosen J. D., Raymond P. A., Saiers J., Shanley J. B., Dittmar T.
 1001 and Stubbins A. (2019) Molecular hysteresis: Hydrologically driven changes in riverine
 1002 dissolved organic matter chemistry during a storm event. *J. Geophys. Res. Biogeosciences*
 1003 **124**, 759–774.

1004 Ward N. D., Keil R. G., Medeiros P. M., Brito D. C., Cunha A. C., Dittmar T., Yager P. L.,
 1005 Krusche A. V. and Richey J. E. (2013) Degradation of terrestrially derived macromolecules
 1006 in the Amazon River. *Nat. Geosci.* **6**, 530–533.

1007 Waterloo M. J., Oliveira S. M., Drucker D. P., Nobre A. D., Cuartas L. A., Hodnett M. G.,
 1008 Langedijk I., Jans W. W. P., Tomasella J., Araujo A. C. de, Pimentel T. P. and Estrada J. C.
 1009 M. (2006) Export of organic carbon in run-off from an Amazonian rainforest blackwater

1010 catchment. *Hydrol. Process.* **20**, 2581–2597.

1011 Webb J. R., Santos I. R., Maher D. T. and Finlay K. (2018) The importance of aquatic carbon
1012 fluxes in net ecosystem carbon budgets: A catchment-scale review. *Ecosystems* **22**, 508–
1013 527.

1014 Zanchi F. B., Meesters A. G. C. A., Waterloo M. J., Kruijt B., Kesselmeier J., Luizão F. J. and
1015 Dolman A. J. (2014) Soil CO₂ exchange in seven pristine Amazonian rain forest sites in
1016 relation to soil temperature. *Agric. For. Meteorol.* **192–193**, 96–107.

1017 Zanchi F. B., Waterloo M. J., Dolman A. J., Groenendijk M., Kesselmeier J., Kruijt B., Bolson
1018 M. A., Luizão F. J. and Manzi A. O. (2011) Influence of drainage status on soil and water
1019 chemistry, litter decomposition and soil respiration in central Amazonian forests on sandy
1020 soils. *Rev. Ambient. e Agua* **6**, 6–29.

1021 Zanchi F. B., Waterloo M. J., Tapia A. P., Alvarado Barrientos M. S., Bolson M. A., Luizão F.
1022 J., Manzi A. O. and Dolman A. J. (2015) Water balance, nutrient and carbon export from a
1023 heath forest catchment in central Amazonia, Brazil. *Hydrol. Process.* **29**, 3633–3648.

1024 Zarnetske J. P., Bouda M., Abbott B. W., Saiers J. and Raymond P. A. (2018) Generality of
1025 hydrologic transport limitation of watershed organic Carbon flux across ecoregions of the
1026 United States. *Geophys. Res. Lett.* **45**, 11702–11711.

1027 Zavarzina A. G., Lisov A. V. and Leontievsky A. A. (2018) The role of ligninolytic enzymes
1028 laccase and a versatile peroxidase of the white-rot fungus *Lentinus tigrinus* in
1029 biotransformation of soil humic matter: Comparative in vivo study. *J. Geophys. Res.*
1030 *Biogeosciences* **123**, 2727–2742.

1031 Zhang X.-P., Yang Z.-L., Niu G.-Y. and Wang X.-Y. (2009) Stable water isotope simulation in
1032 different reservoirs of Manaus, Brazil, by Community Land Model incorporating stable

isotopic effect. *Int. J. Climatol.* **29**, 619–628.

Captions

Figure 1. Detail of the lower Rio Negro catchment northwest of Manaus showing sampling sites of available FT-MS studies. We sampled at two locations north of Manaus (green dots; “Cuieiras”, “Campina”). Two other FT-MS datasets from blackwater sampling locations are shown, including the Rio Negro and connected lakes (black dots; Gonsior et al., 2016) and the lower reach of the Rio Negro and two of its tributaries close to Manaus (red dots; Simon et al., 2019). We accessed watershed limits and river data as shapefiles from www.ore-hybam.org (Seyler et al., 2009). The 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. Differences in the water and SPE-DOM properties. Variables shown are a) pH, b) electrical conductivity, c) dissolved organic carbon concentration, d) difference in $\delta^{18}\text{O}$ values of water compared to regional average precipitation (OIPC estimate), e) mean calendar age of DOC in solid-phase extracts. Samples are grouped into biogeochemical environments (upland Campina; riparian valley, and plateau). Symbols denote sample type (squares, piezometer; star, deep well; circle, stream); open symbols mark “intermediate” valley samples (see main text).

Figure 3. Multivariate analysis based on molecular DOM data. a) Separation of samples in a principal coordinate analysis (PCoA) based on Bray Curtis dissimilarity. The plot shows only the first two coordinates, the third coordinate (5% of explained variability, not shown) did not contribute further to separation. Percentages denote the degree of explained variability in DOM molecular composition. Samples are grouped into biogeochemical environments (“Upland”,

1056 “Valley”, “Plateau”). Symbols denote sample type (square, piezometer; star, deep well; circle,
 1057 stream); open symbols mark “intermediate” valley samples (see main text). b - d) Post-ordination
 1058 gradient fit (function envfit of R package vegan, at 999 permutations) of three different sets of
 1059 variables based on PCoA separation (same as in a). Variables sets are b) DOM indices, c)
 1060 Molecular groups, and d) Formula classes. Significant correlations (Pearson, $p < 0.05$) with the
 1061 ordination are shown as arrows. Arrow length corresponds to the strength of correlation and
 1062 arrows head into the direction of the steepest increase of the respective variable, based on the
 1063 ordination pattern of samples. Variable abbreviations in b): C, H, O, N, S (average numbers of
 1064 respective atoms per formula), MW (molecular weight as mass to charge-ratio), AImod
 1065 (Aromaticity index), DBE, DBEmO, DBE.C (double bond equivalents, DBE minus oxygen,
 1066 DBE/C ratio), H/C (atomic ratio of hydrogen to oxygen, “saturation axis”), O/C (atomic ratio of
 1067 oxygen to carbon, “oxidation axis”), NOSC (nominal oxidation state of carbons), Np, Np%,
 1068 Npi% (number, percentage and relative abundance of N-containing peaks), Sp, Sp%, Spi% (same
 1069 for S-containing peaks). Abbreviations in c): BC (Polycyclic, condensed aromatics, such as
 1070 “Black Carbon”), PP (polyphenols), HU (highly unsaturated), PEP (unsaturated, O- and N-
 1071 containing, such as peptides), UA (unsaturated aliphatics), SUG (very high O content, such as
 1072 sugars). Prefix to PP, HU & UA: op ($O/C \leq 0.5$); or ($O/C > 0.5$). Prefix to BC and PP: lw (< 15
 1073 C atoms), hw (≥ 15 C atoms). Abbreviations in d): CHO (average number of molecular formulae
 1074 containing only C, H and O atoms), CHNO, CHN₂O, CHOS, CHNOS (formulae containing one
 1075 N, two N atoms, one S atom, or both one N and S atom).
 1076 Figure 4. Subsets of molecular formulae showing significant “enrichment” (higher ion
 1077 abundance), i.e., ecosystem specificity, in the plateau (panels a-c), valley WSE (d-f) and upland
 1078 WSE samples (g-i). General WSE (valley and upland) markers and non-significant signals

(common to all samples) are shown in Supporting Information Figure S4. Left column panels (a, d, g) show the average (and max) mass spectrum of each ecosystem. Mid column panels (b, e, h) show the formula subsets in Van Krevelen space (each dot represents a molecular formula defined by its atomic ratios of hydrogen, H/C, and oxygen to carbon, O/C). Formulae are colored according to classes (see legend and numbers of formulae in panels to the right). The plot is divided by solid lines that mark molecular group categories (see also caption of Figure 3, and Supporting Information Table S2). Right column panels (c, f, i) show the same formula subsets from mid-row panels in H/C vs. m/z (mass-to-charge) space. Dotted grey lines are for visual guidance and comparison (at $m/z = 450$, $O/C = 0.5$, and $H/C = 1$). Left- and right-column plots share the m/z abscissa (x-axis) while mid- and right-column plots share the H/C ordinate (y-axis).

Figure 5. Overlap of WSE-specific formulae (significantly enriched formulae) with two independent sets of Rio Negro DOM markers (a-c: data from Gonsior et al., 2016; d-f: data from Simon et al., 2019). Left panels (a, d) show the overlap of three sets of markers (Rio Negro, Valley, Upland) in a Venn diagram created with package “eulerr” in RStudio. Overlap indicates common formulae, i.e., shared information. Areas are scaled to number of formulae, which are given for each subset (note color, area may be dissected). Overlap is shown only for CHO formulae because Rio Negro markers were mainly CHO-type, other formula classes showed no distinct overlap. Right panels show Rio Negro markers of each study (grey filled squares) in chemical space as a Van Krevelen plot (b, e) or H/C vs. m/z plot (c, f; similar visualization as in Figure 4). Colored symbols show the match between datasets (“overlap”) and refer to the bold ellipse in panels a and d, respectively. Plateau markers are additionally added for comparison. Red squares denote the small set of WSE markers that were found to match with both sets of Rio

1102 Negro markers. They were uniquely enriched in upland ($n = 6$) or common to upland and valley
1103 WSE's ($n = 3$). Based on the chemistry of these nine formulae, the black box denotes the wider
1104 area of consistent matching in terms of chemical space (H/C 0.4 – 0.8; O/C 0.4 – 0.8).
1105 Independent of the underlying set of Rio Negro markers, the matching rate was higher for upland
1106 WSE markers (lilac).

Supplementary Material

Molecular links between whitesand ecosystems and blackwater formation in the Rio Negro watershed

C. Simon^{(1),(2)}, T. P. Pimentel⁽³⁾, M. T. F. Monteiro⁽³⁾, L. A. Candido⁽³⁾, D. Gastmans⁽⁴⁾, H. Geilmann⁽⁵⁾, R. da Costa Oliveira⁽³⁾, J. B. Rocha⁽³⁾, E. Pires⁽³⁾, C. A. Quesada⁽³⁾, B. R. Forsberg^{(3),(6)}, S. J. F. Feirrera⁽³⁾, H. B. da Cunha⁽³⁾, G. Gleixner^{*(1)}

⁽¹⁾ Molecular Biogeochemistry, Max Planck Institute for Biogeochemistry (MPI-BGC), Hans Knöll-Str. 10, 07745 Jena, Germany.

⁽²⁾ present address: Institute of Biogeochemistry and Pollutant Dynamics (IBP), ETH Zürich, Universitätstrasse 16, 8092 Zürich, Switzerland.

⁽³⁾ Coordenação de Dinâmica Ambiental (CODAM), Instituto Nacional de Pesquisas da Amazônia (INPA), Av. Efigênio Sales 2239, Aleixo, Manaus, Brazil.

⁽⁴⁾ São Paulo State University (UNESP), Centro de Estudos Ambientais, Av. 24A, 1515, Bela Vista, Rio Claro, São Paulo, Brazil.

⁽⁵⁾ Stable Isotope Laboratory (BGC-IsoLab), Max Planck Institute for Biogeochemistry, Hans Knöll-Str. 10, 07745 Jena, Germany.

⁽⁶⁾ present address: Vermont Agricultural and Environmental Laboratory, 163 Admin Dr, Randolph Center, 05602, Vermont, United States.

*Correspondence: gerd.gleixner@bgc-jena.mpg.de (Gerd Gleixner)

Contents

Introduction.....	3
Supplementary Table S1. Combined data of samples described in this study.....	4
Supplementary Table S2. Molecular indices calculated from FT-MS data.....	5
Supplementary Table S3. Analysis of structure suggestions of potential white-sand markers	6
Supplementary Figure S1. Schematic landscape sections of the studied white-sand ecosystems..	7
Supplementary Figure S2. Water isotope data from this study	8
Supplementary Figure S3. Water isotope data of INPA's meteorological station Reserva Ducke	9
Supplementary Figure S4. Subsets of non-informative molecular formulae.....	10
Supplementary Figure S5. Overview of all subsets of molecular formulae	11
Supplementary Figure S6. Venn diagram of subsets of samples: compound classes.....	12
Supplementary Figure S7. Venn diagrams of subsets of samples: molecular groups	12

31	Supplementary Figure S8. Venn diagram of the three FT-MS datasets used for comparison.....	13
32	Supplementary Figure S9. Cluster analysis of the three FT-MS datasets used for comparison...	13
33	Supplementary Text S1. General comparison of published Rio Negro DOM data sets.....	14
34	Supplementary Material References	16
35		
36		

Introduction

This Supporting Information file contains one supporting text resource (Text S1), nine supporting figures (Figures S1 to S9), three supporting tables (Table S1 to S3), and information on three supporting data sets (Data Sets S1, S2, and S3). These data sets are openly available from <https://doi.org/10.1594/PANGAEA.922606>. This Supporting Information file contains 23 references.

The text resource “Text S1” presents an additional comparative analysis of the three datasets that were part of the main study (→ interlinked with Figures S8 and S9 and Data Set S2).

“Figure S1” shows schematic landscape section of the two sample gradients. “Figure S2” presents water isotope data in a common $\delta^{18}\text{O}$ - $\delta^2\text{H}$ plot, along with available local meteoric water lines, rain data, and OIPC averages. “Figure S3” presents monthly rain water isotope composition from INPA’s meteorological station in Reserva Ducke, Manaus, Brazil. “Figure S4” presents molecular data of subsets of non-indicative molecular formulae (compare to Figure 4 of main text). “Figure S5” present data for all indicative and non-indicative subsets of formulae for a general comparison of diversity, molecular groups, formula classes, and DOM indices. “Figure S6” and “Figure S7” show the overlap in formulae between ecosystems in terms of compound class and molecular group. “Figure S8” shows the result of a Venn analysis of the three datasets (see “Text S1”), and “Figure S9” shows a related cluster analysis for general similarity among individual samples of all three datasets.

“Table S1” includes all the environmental data obtained for each sample. “Table S2” is a compilation of the derived DOM indices that were used for chemical description of DOM samples with references for each index. “Table S3” summarizes the structural information obtained for nine potential whitesand Rio Negro markers from PubChem.

“Data Set S1”, an .xlsx file, contains the crosstab of all molecular formulae used for the analyses throughout the main manuscript, the DOM index data, ecosystem averages, ecosystem fingerprint assignments, Rio Negro marker overlap, and evaluation of structural data from PubChem. “Data Set S2”, also an .xlsx file, contains the merged crosstab that was used for a general dataset comparison of whitesand DOM and openly available Rio Negro datasets. “Data Set S3” is a .docx file containing the list of structure suggestions for nine potential whitesand Rio Negro markers from PubChem, including the structural formulae (which are not provided in “Data Set S1”).

Data sets are available from the Pangaea Data Publisher via the following link:
<https://doi.org/10.1594/PANGAEA.922606>

Supplementary Table S1. Combined data of samples described in this study. All samples were taken in 2017. Greyed entries denote problematic data (see additional comments below table). Abbreviations: EC, electrical conductivity; F14C, Fraction Modern; $\Delta^{14}\text{C}$, correction accounting for decay between sample collection and measurement; EE, extraction efficiency based on DOC of samples and SPE extracts.

ID	Group	Specifier	Date	Type	Depth* [m]	pH	EC [$\mu\text{S}/\text{cm}$]	DOC [mg/L]	$\delta^2\text{H}$ [‰]	$\delta^{18}\text{O}$ [‰]	d-excess [‰]##	F ¹⁴ C	$\Delta^{14}\text{C}$ [‰]	Cal. Age [years]	EE [%]
PR11	Plateau	-	11/01	Piez.	0.96*	4.2	14	0.89	-26.8	-5.03	13.44	n.d.**	-	-	6
PR10	Valley	Intermediate	10/31	Piez.	n.d.	4.2	16	2.31	-24.2	-4.50	11.8	0.974 [#]	-33.6 [#]	(2026) [#]	63
PR9	Valley	Upland-like	10/31	Piez.	0.24*	3.6	54	37.0	-20.2	-3.96	11.48	1.055	46.8	2009	71
PR8	Valley	Upland-like	10/31	Piez.	0.15*	3.6	50	34.1	-16.6	-3.61	12.28	1.056	47.0	2009	80
PR7	Valley	Upland-like	10/31	Piez.	1.12*	3.6	50	37.5	-29.8	-5.11	11.08	1.055	46.1	2009	72
PT6	Valley	Intermediate	10/31	Piez.	n.d.	3.9	29	31.3	-26.4	-4.61	10.48	1.021	12.3	2016	33
PR6	Plateau	-	11/01	Piez.	1.86*	4.5	11	1.80	-24.6	-4.84	14.12	n.d.**	-	-	22
PP1	Plateau	-	11/01	Well	39.0	4.5	12	0.54	-26.6	-4.96	13.08	n.d.**	-	-	42
PP2	Plateau	-	11/01	Well	35.0	4.7	10	0.56	-28.9	-4.86	9.98	n.d.**	-	-	20
RA	Valley	Intermediate	10/31	River	0	4.3	14	6.29	-23.3	-4.44	12.22	1.035	26.4	2013	80
P2	Upland	-	11/02	Piez.	2.4	3.6	55	36.9	-7.2	-2.75	14.8	1.063	53.8	2008	60
P4	Upland	-	11/02	Piez.	1.5	3.9	40	28.2	-18.2	-3.77	11.96	1.055	46.4	2009	72
P5	Upland	-	11/02	Piez.	1.5	3.8	43	30.4	-15.9	-3.69	13.62	1.072	63.6	2006	73
P6	Upland	-	11/02	Piez.	1.5	3.6	50	45.7	-15.9	-3.67	13.46	1.079	70.5	2004	63
P7	Upland	-	11/02	Piez.	1.5	3.6	52	38.5	-17.6	-3.77	12.56	1.074	64.7	2005	79
RC	Upland	-	11/02	River	0	3.6	54	46.4	-16.4	-3.79	13.92	1.070	61.7	2006	62

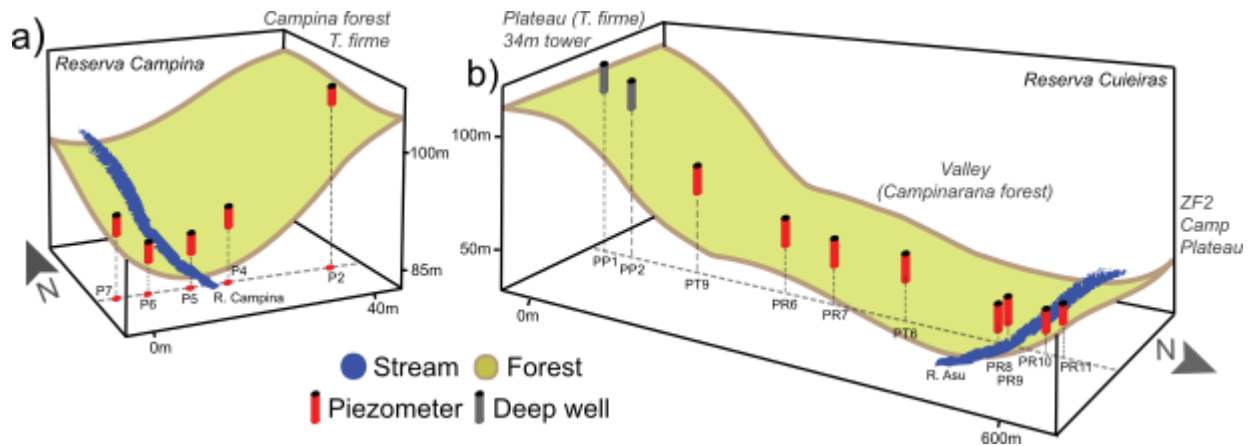
* Reserva Cuieiras: In piezometers, water level below the surface at sampling (daily mean, hourly data), in wells: maximum depth. Reserva Campina: piezometers, max. depth. ** n.d., not determined due to the limited amount of extract. [#] Value likely influenced by ¹⁴C-dead contaminant signal. ^{##} Calculated based on the formula d-excess = $\delta^2\text{H} - 8 * \delta^{18}\text{O}$ (Dansgaard, 1964).

79 **Supplementary Table S2.** Molecular indices calculated from FT-MS data.

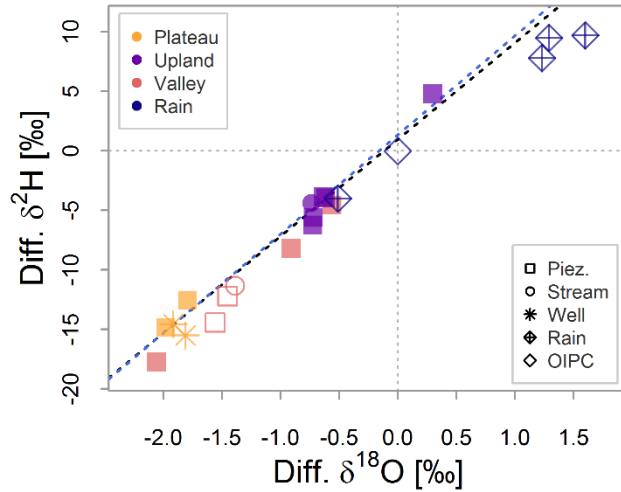
Index	Explanation	Calculation/ definition	Reference of use
H/C	Atomic ratio of hydrogen to carbon in a molecular formula	H/C	Kew et al., 2017; Kim et al., 2003
O/C	Atomic ratio of oxygen to carbon in a molecular formula	O/C	
DBE	Double Bond Equivalents	$1+0.5*(2*C-H+N+P)$	Koch and Dittmar, 2016, 2006
Almod	Aromaticity index	$[1+C-0.5*O-S-0.5*(N+P+H)]/C-0.5*O-N-S-P$	
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	Herzprung et al., 2014; Raeke et al., 2017; Roth et al., 2013
NOSC	Nominal Oxidation State of Carbon	$4-[(4*C+1*H-3*N-2*O-2*S)/C]$	Riedel et al., 2013; see also Boye et al., 2017; Kroll et al., 2011
CHO	Formulae containing only oxygen besides C and H	Count formulae	Many classes used, e.g. Pomerantz et al., 2011; Zhurov et al., 2013
CHNO	Formulae containing additional nitrogen		
CHOS	Formulae containing additional sulfur		
CHNOS	Formulae containing nitrogen and sulfur		
BC	Polycyclic, condensed aromates, such as “Black Carbon”	$Almod \geq 0.66$	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	$0.5 \geq Almod < 0.66$	
HU	Highly unsaturated compounds	$Almod \geq 0.5$; $H/C < 1.5$; $O/C < 0.9$	
UA	Unsaturated aliphatics	$1.5 \geq H/C < 2$; $O/C < 0.9$; $N = 0$	
PEP	Unsaturated, O- and N-containing compound, such as peptides	$1.5 \geq H/C < 2$; $O/C < 0.9$; $N > 0$	
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	$O/C \geq 0.9$	
Prefix (BC, PP)	„lw“ – very low molec. weight, „hw“ – higher molec. weight	Additional constraint: $C < 15$ or ≥ 15	
Prefix (PP, HU, UA)	„or“ – rich in oxygen, „op“ – poor in oxygen	Additional constraint: $O/C > 0.5$ or ≤ 0.5	

Supplementary Table S3. Structural features based on PubChem (search conducted on 28th May 2020) for nine specific Rio Negro markers (see Figure 5) also found in WSE-DOM. Average = Sum of respective feature across structures/ Number of structures. Row color gradients indicate low (green) to high (red) values. WSE, marker common to upland and valley WSEs; UPL, marker enriched in upland WSE; w, with; w/o, without.

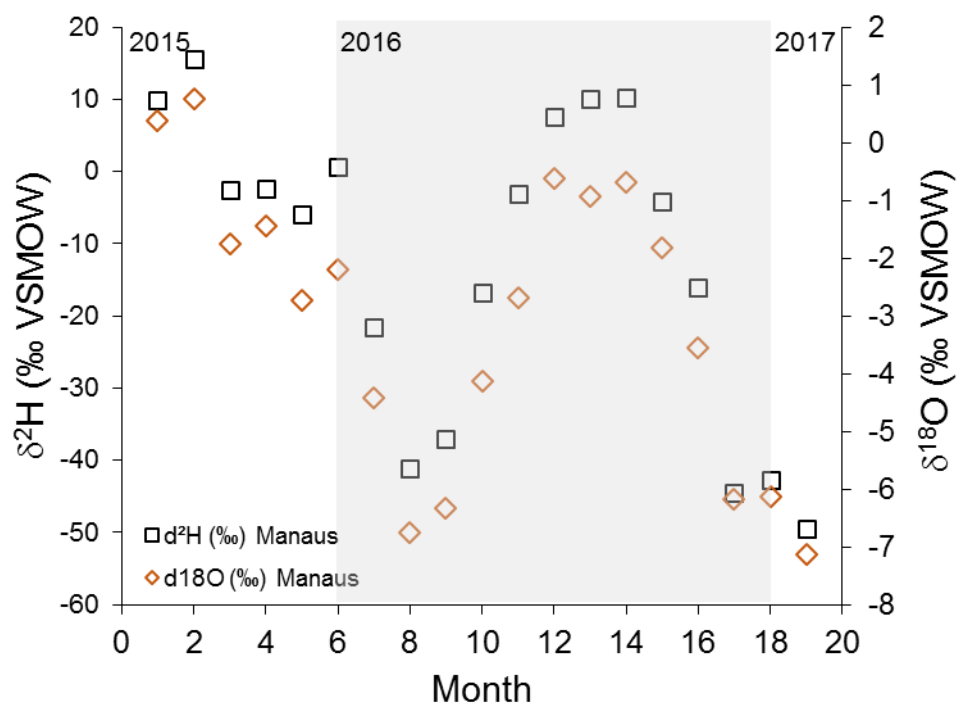
Formula	C ₁₀ H ₆ O ₆	C ₁₁ H ₆ O ₆	C ₁₁ H ₆ O ₇	C ₁₁ H ₆ O ₈	C ₁₂ H ₁₀ O ₇	C ₁₇ H ₈ O ₇	C ₂₁ H ₁₂ O ₈	C ₂₂ H ₁₄ O ₉	C ₂₂ H ₁₄ O ₉
Ecosystem	UPL	WSE	UPL	UPL	WSE	UPL	WSE	UPL	UPL
Indices based on molecular formula									
Molecular group	BC	BC	BC	BC	PP	BC	BC	PP	PP
Mass [Da]	221	233	249	264	265	323	391	421	429
DBE	8	9	9	9	8	14	16	16	14
Al _{MOD}	0.71	0.75	0.73	0.71	0.53	0.78	0.71	0.66	0.59
O/C	0.60	0.55	0.64	0.73	0.58	0.41	0.38	0.41	0.55
H/C	0.60	0.55	0.55	0.55	0.83	0.47	0.57	0.64	0.70
Hits (PubChem)	108	40	16	5	96	20	18	18	7
PubChem: Fused rings									
Average per molecule	1.88	2.28	1.69	1.40	1.58	2.45	3.28	2.61	1.71
Hits > one ring [%]	76	88	63	40	52	100	94	56	43
Hits > two rings [%]	11	40	6	0	6	40	61	44	14
PubChem: Aromatic rings									
Average per molecule	0.92	0.95	1.06	0.60	0.96	1.80	2.39	2.56	2.14
Hits > one ring [%]	8	5	6	20	14	80	100	100	100
Hits > two rings [%]	0	0	0	0	0	5	33	56	14
PubChem: Quinone-like rings									
Average per molecule	0.24	0.05	0.19	0	0.18	0.05	0.17	0.06	0
Total hits [%]	17	5	19	0	10	5	11	6	0
PubChem: Oxygen heterocycles									
Average per molecule	0.87	1.35	0.69	1.20	0.64	1.65	2.06	0.89	0.43
Hits > zero rings [%]	68	90	56	80	52	95	89	50	29
Hits > one ring [%]	19	45	13	40	10	70	72	17	14
Hits w pyran-like ring(s) [%]	29	45	25	40	20	15	50	28	0
Hits w furan-like ring(s) [%]	30	50	19	20	18	70	39	17	14
Hits w other O-heterocycle(s) [%]	13	28	19	20	17	20	22	11	29
PubChem: Specific scaffolds/ substructures									
Hits w chromene unit(s) [%]	13	0	19	0	10	0	11	0	0
Hits w naphthalene unit(s) [%]	26	48	19	0	16	10	28	11	0
Hits w benzoic acid/ phenol unit(s) [%]	16	3	31	20	32	0	28	67	100
Hits w benzofuran unit(s) [%]	14	20	13	0	5	70	22	22	29
PubChem: Functional groups (average per molecule)									
Double bonds	3.23	3.55	3.50	3.20	3.61	6.05	8.39	8.39	6.57
Carboxyl (COOH)	0.72	0.80	1.06	1.40	0.97	0.20	0.28	1.61	1.86
Hydroxyl (OH), w/o carboxyl	1.21	0.33	0.81	0.60	1.58	0.60	2.11	1.89	2.00
Carbonyl (C=O), w/o carboxyl	2.02	2.25	2.63	2.60	1.61	3.90	2.56	1.94	3.00
Methyl (Me), w/o methoxy	0.01	0.15	0.06	0	0.25	0	0.06	0.28	0.71
Methoxy (MeO)	0.05	0.08	0.06	0	0.48	0	0.22	0.11	0.29
Lactone (carboxyl ring condensation)	0.57	0.80	0.31	0.80	0.28	1.60	0.67	0.11	0.71
Ether (COC), w/o methoxy & lactone	0.65	0.83	0.81	0.80	0.96	0.40	1.56	1.56	1.14



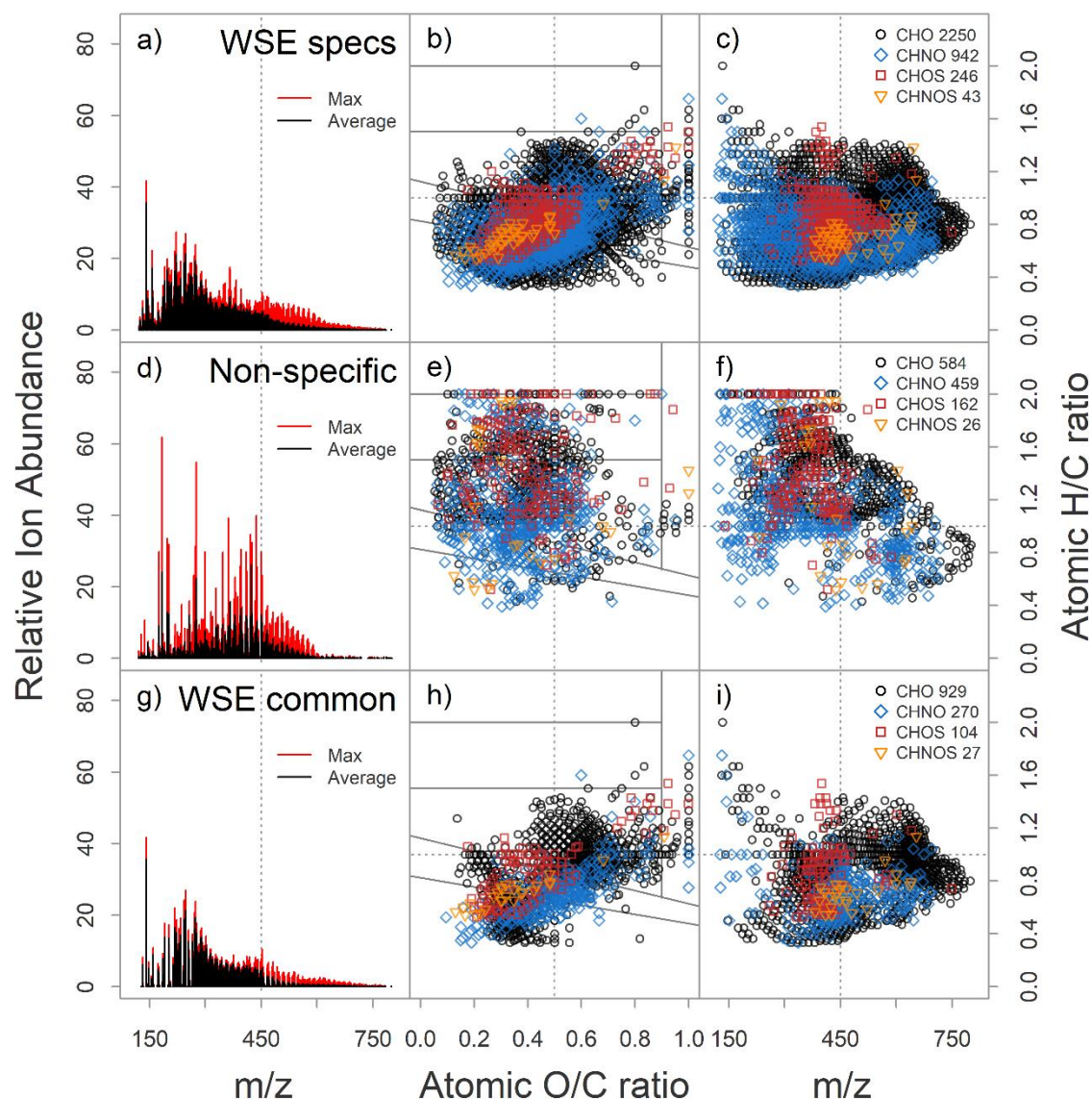
Supplementary Figure S1. Schematic landscape sections of the two sampled whitesand ecosystems with sample locations along transects. a) Upland WSE Campina forest transect at Reserva Campina. b) Elevated plateau with intersected riparian valley WSE at Reserva Cuieiras. Note differences in scale.



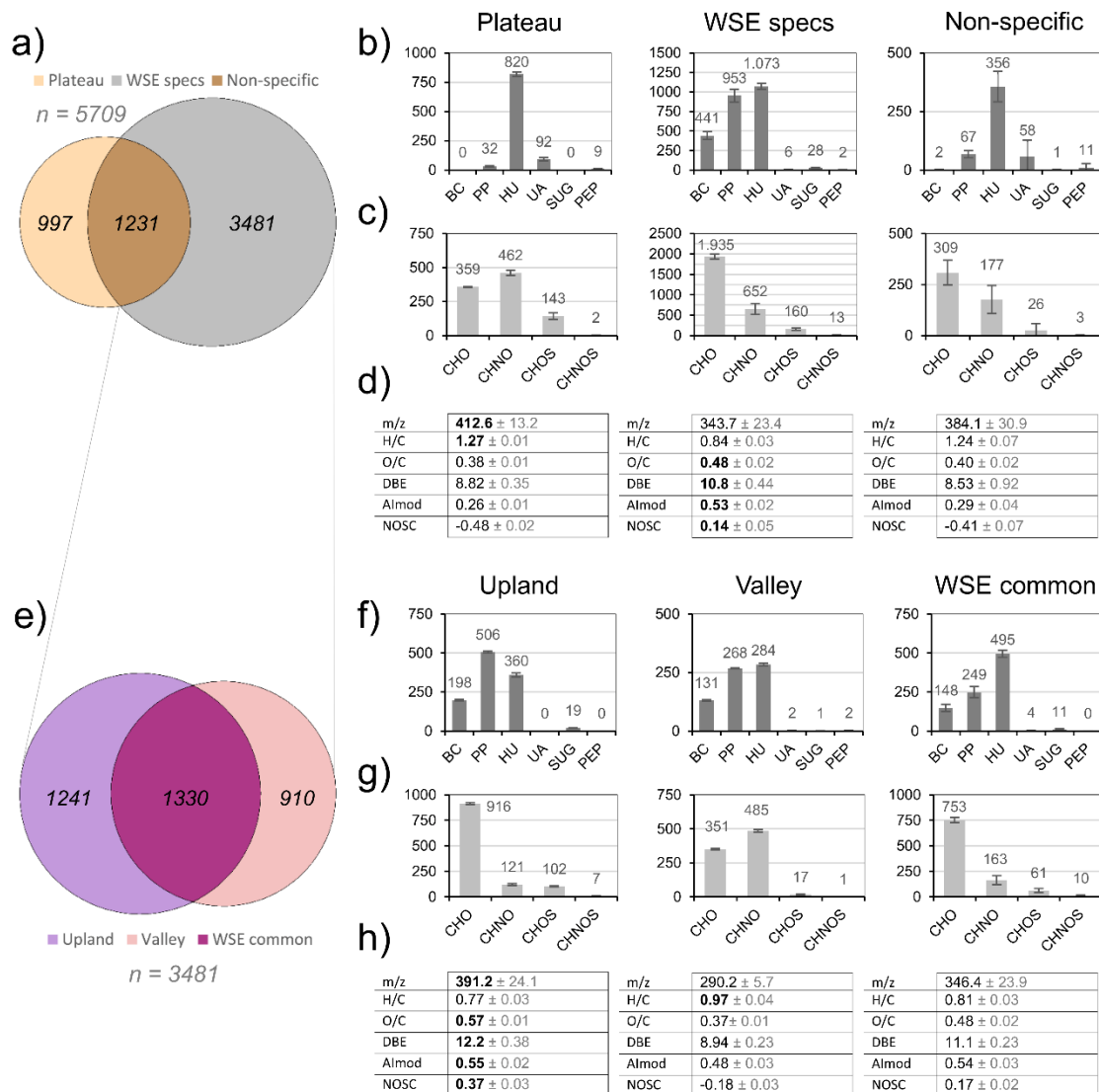
Supplementary Figure S2. Water isotope data in relative delta notation against predicted average precipitation by OIPC (blue open diamond at plot origin; OIPC 3.0 estimate, see methods) in water samples taken in October/ November 2017. Black dotted line: Local meteoric water line (LMWL) constructed from data collected monthly by the INPA climatology station, located in the Adolpho Ducke Forest Reserve, Manaus, Brazil (data 08/15 - 02/17): $\delta^2\text{H} = 8.343 \cdot \delta^{18}\text{O} + 13.362$ ($r = 0.99$, $n = 19$). Four selected rain datapoints from October and November 2015 and 2016 are shown as blue crossed diamonds. Blue dotted line: LMWL constructed from monthly $\delta^2\text{H}$ and $\delta^{18}\text{O}$ 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: $\delta^2\text{H} = 8.14 \cdot \delta^{18}\text{O} + 12.96$ ($r = 0.98$, $n = 186$).



Supplementary Figure S3. Water isotope data from INPA's meteorological station at Reserva Ducke, north of Manaus, over the course of the year (data coverage: 08/15 – 02/17; n=19). Grey shading marks 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.

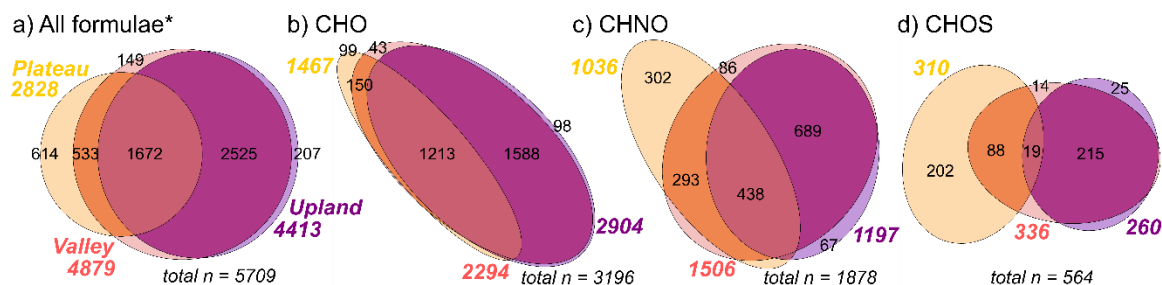


Supplementary Figure S4. Subsets of non-informative molecular formulae (not related uniquely to one of the three biogeochemical environments but common to two or three). Left column panels (a, d, and g) show the average (and max) mass spectrum of each sample set. Mid column panels (b, e, and h) show the formula subsets in chemical space (Van Krevelen plot): each molecular formula is represented by a dot according to its atomic ratios of hydrogen (H/C) and oxygen to carbon (O/C; see additional grouping into formula classes and respective numbers of formulae in legends. See also Figure 4 in the main text. Top row shows dominant formulae in whitesand ecosystem samples (“WSE specs”) as opposed to the plateau environment. The middle row shows common formulae (i.e., non-significant differences in ion abundance across all three biogeochemical environments; “Non-specific”). The bottom row plots showing formulae shared between whitesand ecosystem samples (“WSE common”).

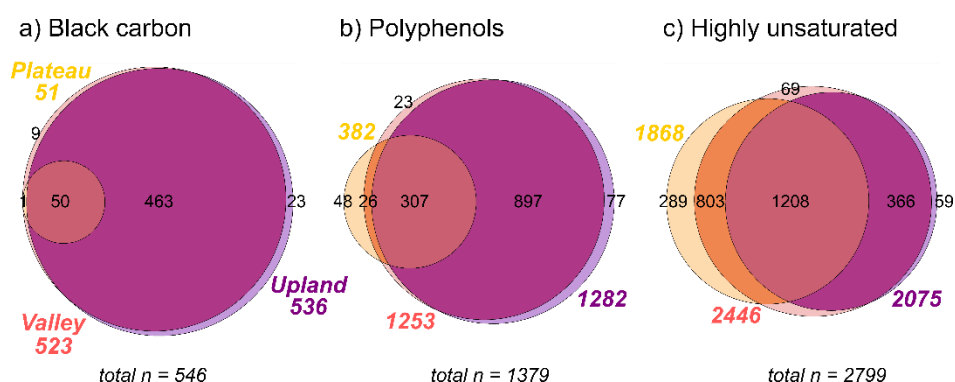


124

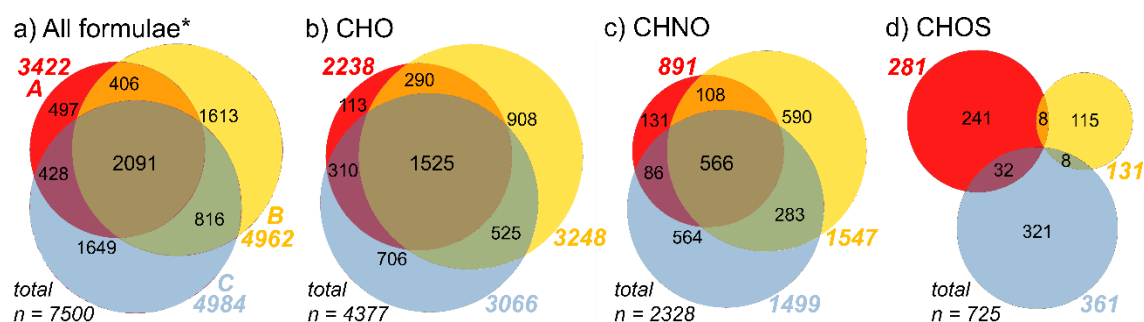
125 **Supplementary Figure S5.** Overview of all subsets of molecular formulae, showing results of
 126 the comparison among samples from all three biogeochemical environments (a – d) and the WSE
 127 sites (Upland and Valley samples) only (d – h). Names of subsets relate to data shown in Figure
 128 4 and Figure S4. Venn diagrams show significantly (Pearson, $p < 0.05$) enriched formulae of
 129 each subset. Panels b – d (and f – h) show differences among subsets in terms of b/ f) number of
 130 formulae classified into molecular groups (BC, polycyclic, condensed aromates, such as “Black
 131 Carbon”; PP, polyphenols; HU, highly unsaturated; UA, unsaturated aliphatics; SUG, very high
 132 O content, such as sugars; PEP, unsaturated, O- and N-containing, such as peptides), c/ g)
 133 number of formulae classified into formula classes (CHO, formulae containing only C, H and O
 134 atoms; CHNO/ CHOS/ CHNOS, formulae containing one N or two N atoms, one S atom, or both
 135 N and S atoms), and d/ h) DOM indices, based on ion-abundance weighted averages across
 136 samples of each subset (m/z, molecular weight as mass to charge-ratio; H/C, atomic ratio of
 137 hydrogen to oxygen; O/C, atomic ratio of oxygen to carbon; DBE, double bond equivalents;
 138 Almod, Aromaticity index; NOSC, nominal oxidation state of carbons).



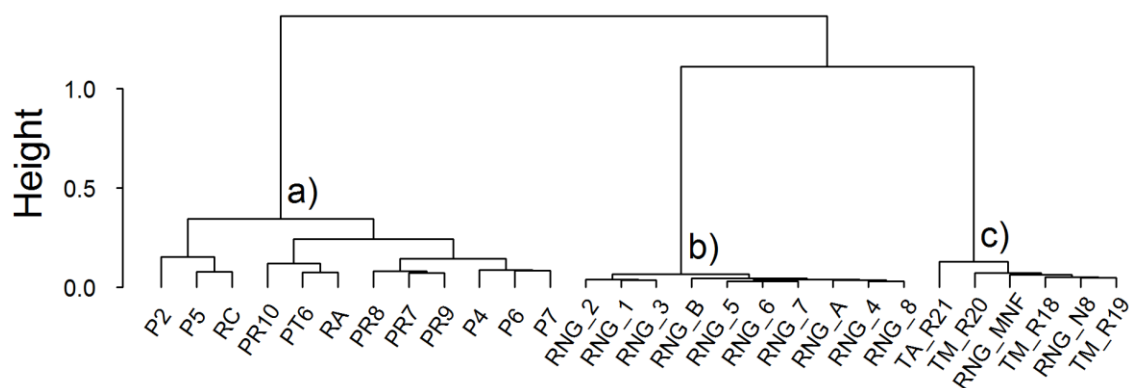
Supplementary Figure S6. Venn diagrams showing the overlap between molecular formulae between subsets of DOM samples, for a) all formulae (*including CHNOS formulae), b) formulae containing only carbon, hydrogen and oxygen atoms, c) formulae containing one or two N atoms, and d) formulae containing an S atom. Overlap between sample sets is highest in CHO formulae and lowest in CHNO and CHOS formulae.



Supplementary Figure S7. Venn diagrams showing the overlap between molecular formulae between subsets of DOM samples, for a) formulae classified as “black-carbon”-like, b) polyphenol-like, and “highly unsaturated”. For molecular group definitions, see Table S2. Valley and upland samples are highly similar in terms of formulae present, but their intensity differs (see Figure 4, and Figures s4 and S5). The plateau samples are poor in black-carbon and polyphenol-like formulae but are similarly rich in “highly unsaturated” compounds.



Supplementary Figure S8. Venn diagram showing the overlap in terms of molecular formulae in the three different FT-MS 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: 70 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.



Supplementary Figure S9. Result of the cluster analysis taking into account all formulae present in each measurement of the three datasets under study across the mass range m/z 180 – 800 (number of formulae = 7500). 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. Groundwater/ headstream DOM from whitesand areas (a) is most dissimilar from river DOM (b, c). Spatial variability among sample sets is more pronounced in the groundwater dataset (a), probably due to lower heterogeneity in aquatic settings (mixing, etc.). However, even headwater streams differ strongly (RA, RC) from river samples (b, c), hence suggesting compositional changes during downstream transport.

Supplementary Text S1. General comparison of published Rio Negro DOM data sets.

Data from blackwater samples of the two datasets (Gonsior et al., 2016; Simon et al., 2019) 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 μ l/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). Besides site and lab effects, accumulation time has to be regarded as the main factor of variation under these otherwise similar measurement conditions (Hawkes et al., 2016; Simon et al., 2018).

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. To bring datasets into comparable format, we removed 1) formulae detected below m/z 180 and above m/z 800 (to account for different scan ranges), 2) formulae containing P atoms or two S atoms or three to four N atoms (32 formulae with N₃ excluded in Gonsior et al. 2016; 57 P, two S₂ and twelve N₃₋₄ formulae excluded in Simon et al. 2019). The remaining lists of each dataset were then merged by molecular formula and compared by Venn diagrams (overlap in terms of formula populations). 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”).

The merged master list contained 7500 molecular formulae that represent an updated inventory of the Rio Negro watershed DOM spectrum (Data Set S2). A Venn analysis of the whole dataset revealed a common set of 2091 formulae (Figure S8) 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%). The similarity in CHO formulae was a little higher compared to the total set of formulae (35%; 16%; 21%; 3%) whereas in terms of the CHNO formulae, similarity was lower than based on the total set (24%; 24%; 25%; 5%). 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 (whitesand data set).

The comparison of datasets revealed that nitrogen- and especially sulfur containing formulae clearly differentiate the three datasets considered in this study. 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; Peyton Smith et al., 2017; Lynch et al., 2019), 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. Sulfur-containing formulae may also originate from contamination, by e.g. sulfonic acids.

The 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 S9). In general, the three datasets were clearly separated, and river samples were more similar to another than to any type of WSE-DOM. 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. Samples from a suite of channels separated by river islands (cluster b; Gonsior et al. 2016), were astonishingly similar to each other besides the large spatial extent covered, as compared to Rio Negro and tributary samples (cluster c; Simon et al. 2019) and headwaters (cluster a, this study) that were more dissimilar. This likely reflects a more homogenized aquatic DOM pool as compared to soil environments (Kellerman et al., 2015; Lynch et al., 2019).

Supplementary Material References

- Boye K., Noël V., Tfaily M. M., Bone S. E., Williams K. H., Bargar J. R. and Fendorf S. (2017) Thermodynamically controlled preservation of organic carbon in floodplains. *Nat. Geosci.* **10**, 415–419.
- D’Andrilli J., Cooper W. T., Foreman C. M. and Marshall A. G. (2015) An ultrahigh-resolution mass spectrometry index to estimate natural organic matter lability. *Rapid Commun. Mass Spectrom.* **29**, 2385–2401.
- Dansgaard W. (1964) Stable isotopes in precipitation. *Tellus* **16**, 436–468.
- Gonsior M., Valle J., Schmitt-Kopplin P., Hertkorn N., Bastviken D., Luek J., Harir M., Bastos W. and Enrich-Prast A. (2016) Chemodiversity of dissolved organic matter in the Amazon Basin. *Biogeosciences* **13**, 4279–4290.
- Hawkes J. A., Dittmar T., Patriarca C., Tranvik L. and Bergquist J. (2016) Evaluation of the Orbitrap Mass Spectrometer for the Molecular Fingerprinting Analysis of Natural Dissolved Organic Matter. *Anal. Chem.* **88**, 7698–7704.
- Herzsprung P., Hertkorn N., von Tümpling W., Harir M., Frieze K. and 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.
- Kellerman A. M., Dittmar T., Kothawala D. N. and Tranvik L. J. (2014) Chemodiversity of dissolved organic matter in lakes driven by climate and hydrology. *Nat. Commun.* **5**, 3804.
- Kellerman A. M., Kothawala D. N., Dittmar T. and Tranvik L. J. (2015) Persistence of dissolved organic matter in lakes related to its molecular characteristics. *Nat. Geosci.* **8**.

252 Kew W., Blackburn J. W. T., Clarke D. J. and Uhrín D. (2017) Interactive van Krevelen diagrams -
 253 Advanced visualisation of mass spectrometry data of complex mixtures. *Rapid Commun. Mass*
 254 *Spectrom.* **31**, 658–662.

255 Kim S., Kramer R. W. and Hatcher P. G. (2003) Graphical Method for Analysis of Ultrahigh-Resolution
 256 Broadband mass spectra of Natural Organic Matter, the Van Krevelen diagram. *Anal. Chem.* **75**,
 257 5336–5344.

258 Koch B. P. and Dittmar T. (2006) From mass to structure: An aromaticity index for high-resolution mass
 259 data of natural organic matter. *Rapid Commun. Mass Spectrom.* **30**, 250.

260 Koch B. P. and Dittmar T. (2016) From mass to structure: An aromaticity index for high-resolution mass
 261 data of natural organic matter. *Rapid Commun. Mass Spectrom.* **30**, 250.

262 Kroll J. H., Donahue N. M., Jimenez J. L., Kessler S. H., Canagaratna M. R., Wilson K. R., Altieri K. E.,
 263 Mazzoleni L. R., Wozniak A. S., Bluhm H., Mysak E. R., Smith J. D., Kolb C. E. and Worsnop D.
 264 R. (2011) Carbon oxidation state as a metric for describing the chemistry of atmospheric organic
 265 aerosol. *Nat. Chem.* **3**, 133–139.

266 Lavonen E. E., Kothawala D. N., Tranvik L. J., Gonsior M., Schmitt-Kopplin P. and Köhler S. J. (2015)
 267 Tracking changes in the optical properties and molecular composition of dissolved organic matter
 268 during drinking water production. *Water Res.* **85**, 286–294.

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

272 Peyton Smith A., Bond-Lamberty B., Benscoter B. W., Tfaily M. M., Hinkle C. R., Liu C. and Bailey V.
 273 L. (2017) Shifts in pore connectivity from precipitation versus groundwater rewetting increases soil
 274 carbon loss after drought. *Nat. Commun.* **8**, 1335.

275 Pomerantz A. E., Mullins O. C., Paul G., Ruzicka J. and Sanders M. (2011) Orbitrap mass spectrometry:
 276 A proposal for routine analysis of nonvolatile components of petroleum. *Energy and Fuels* **25**,
 277 3077–3082.

278 Raeke J., Lechtenfeld O. J., Tittel J., Oosterwoud M. R., Bornmann K. and Reemtsma T. (2017) Linking
 279 the mobilization of dissolved organic matter in catchments and its removal in drinking water
 280 treatment to its molecular characteristics. *Water Res.* **113**, 149–159.

281 Riedel T., Zak D., Biester H. and Dittmar T. (2013) Iron traps terrestrially derived dissolved organic
 282 matter at redox interfaces. *PNAS* **110**, 10101–5.

283 Rossel P. E., Bienhold C., Boetius A. and Dittmar T. (2016) Dissolved organic matter in pore water of
 284 Arctic Ocean sediments: Environmental influence on molecular composition. *Org. Geochem.* **97**,
 285 41–52.

286 Roth V.-N., Dittmar T., Gaupp R. and Gleixner G. (2013) Latitude and pH driven trends in the molecular
 287 composition of DOM across a north south transect along the Yenisei River. *Geochim. Cosmochim.*
 288 *Acta* **123**, 93–105.

289 Šantl-Temkiv T., Finster K., Dittmar T., Hansen B. M., Thyraug R., Nielsen N. W. and Karlson U. G.

290 (2013) Hailstones: A Window into the Microbial and Chemical Inventory of a Storm Cloud. *PLoS*
 291 *One* **8**.

292 Seidel M., Beck M., Riedel T., Waska H., Suryaputra I. G. N. A., Schnetger B., Niggemann J., Simon M.
 293 and Dittmar T. (2014) Biogeochemistry of dissolved organic matter in an anoxic intertidal creek
 294 bank. *Geochim. Cosmochim. Acta* **140**, 418–434.

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

298 Simon C., Roth V.-N., Dittmar T. and Gleixner G. (2018) Molecular Signals of Heterogeneous Terrestrial
 299 Environments Identified in Dissolved Organic Matter: A Comparative Analysis of Orbitrap and Ion
 300 Cyclotron Resonance Mass Spectrometers. *Front. Earth Sci.* **6**, 1–16.

301 Zhang X.-P., Yang Z.-L., Niu G.-Y. and Wang X.-Y. (2009) Stable water isotope simulation in different
 302 reservoirs of Manaus, Brazil, by Community Land Model incorporating stable isotopic effect. *Int. J.*
 303 *Climatol.* **29**, 619–628.

304 Zhurov K. O., Kozhinov A. N. and Tsybin Y. O. (2013) Evaluation of high-field orbitrap fourier
 305 transform mass spectrometer for petroleomics. *Energy & Fuels* **27**, 2974–2983.

306