

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

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

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

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

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

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

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

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

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

17 Highlights

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

23 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

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

52

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

55

56 **1 Introduction**

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

70 Blackwater river basins stand out as hotspots of DOC release in the tropics (Junk et al., 2011).
71 The Rio Negro basin for example, which covers roughly 10% of the area of the Amazon river
72 basin, accounts for an annual DOC export of 5.2 – 6.7 Tg C (Coynel et al., 2005; Guinoiseau et
73 al., 2016) equivalent to 17-23% of the Amazon's total DOC export (Raymond and Spencer,
74 2014). Tropical blackwater rivers such as the Rio Negro are thus pivotal in understanding the
75 global carbon cycle and its response to environmental change (Alvarez-Cobelas et al., 2012;
76 Raymond and Spencer, 2014; Webb et al., 2018).

77 Qualitative analyses of DOC composition have shown great potential to track processes and
78 source contributions on the watershed scale (Creed et al., 2015; Riedel et al., 2016; Hutchins et
79 al., 2017; Drake et al., 2019; Spencer et al., 2019). Processes such as deforestation (Drake et al.,
80 2019; James et al., 2019), drainage (Moore et al., 2013) or warming (Drake et al., 2018a) all
81 affect the molecular composition or age of leached DOC. Progress in qualitative DOC measures
82 could thus effectively complement global and regional modelling efforts based on quantitative
83 DOC export data (Alvarez-Cobelas et al., 2012; Lv et al., 2019). However, we do know
84 surprisingly little about potential molecular markers in DOM and their environmental fate.
85 Traditional targeted approaches aiming at terrestrial source markers, namely lignin phenols, can
86 be questioned due to findings suggesting their fast turnover in soils (Gleixner et al., 2002; Hernes
87 et al., 2007; Marschner et al., 2008) and within the land-to-ocean continuum (Cao et al., 2018),
88 and potential autochthonous sources in marine systems (Powers et al., 2019). Only a few authors
89 have addressed the traceability and stability of ecosystem imprints within the aquatic land-to-
90 ocean continuum, and its conditions, in a non-targeted way by means of ultrahigh resolution
91 mass spectrometry to identify novel markers (FT-MS; Roth et al., 2014; Medeiros et al., 2016;
92 Hutchins et al., 2017; Wagner et al., 2019). This gap is due to the limited availability of these

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

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

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

128 Previous studies that assessed terrestrial sources of Rio Negro DOM largely supported the
129 older hypotheses that WSEs, and more specifically the widely distributed riparian corridors are
130 responsible for the highest amount of carbon export (Junk, 1993; Remington et al., 2007; Melack
131 and Hess, 2010; Bardy et al., 2011). However, this DOC export is mainly controlled by
132 precipitation amount and flooding events (McClain et al., 1997; Remington et al., 2007; Zanchi
133 et al., 2015). This is in line with the finding that in riparian settings, DOC generally shows
134 transport-limited behavior, meaning that its absolute flux scales with discharge (Musolff et al.,
135 2017; Zarnetske et al., 2018). As a result, water passing through the riparian zone continuously
136 leaches existing reserves of processed organic matter (Ledesma et al., 2015; Laudon and
137 Sponseller, 2018; Tiegs et al., 2019). In line with the older hypothesis, McClain and coworkers
138 reported low annual DOC yields for widespread plateau areas (terra firme – ferralsol; 2 g DOC

139 m-2 yr-1) but large yields for a relatively small Campina WSE catchment (40 g DOC m-2 yr-1)
140 in the lower Negro basin, north of Manaus. The authors predicted that a WSE molecular DOM
141 signal would thus be easily detectable in higher order rivers within the Rio Negro catchment
142 (McClain et al., 1997), and later studies conducted in the same region corroborated this
143 hypothesis (Remington et al., 2007). DOM from well-developed podzols reflected best the
144 chemical properties of DOM isolates from local groundwater and nearby rivers, showing the fast
145 transit of DOM in well-drained sandy soils with low sorption potential (Remington et al., 2007;
146 Bardy et al., 2011). However, no subsequent markers were identified or tracked in the lower
147 reaches of the stream network to assess their environmental fate or marker potential (Bardy et al.,
148 2011). Such novel markers are however needed to study variations in DOM export and
149 spatiotemporal dynamics in riverine DOM sources within a catchment (Hutchins et al., 2017;
150 Bernhardt et al., 2018; Laudon and Sponseller, 2018).

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

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

166 **2 Materials and Methods**

167 *2.1 Field sites and sampling procedures*

168 Soil water samples were taken in early November 2017 at the onset of the rainy season in two
169 protected forest reserves under the responsibility of the Instituto Nacional de Pesquisas da
170 Amazônia (INPA) in Manaus, Brazil (Figure 1; Supporting Information Figure S1). Both
171 reserves, the Reserva Biológica do Cuieiras – ZF2 (2°36'32.67" S, 60°12'33.48" W, at 40–110
172 m above sea level) and the Reserva Biológica de Campina (2°35'30.26" S, 60°01'48.79" W, at
173 93–101 m a.s.l.) are located about 60 - 70 km north of Manaus (Zanchi et al., 2014; Marques et
174 al., 2016). We conducted sampling from 31st October – 2nd of November, and all lab procedures
175 followed within three days. The geological setting, landscape structure, forest composition, and
176 soil characteristics have been described in detail by Zanchi et al. (2014). Broad swampy valleys
177 surrounded by elevated plateaus cover about half of the Cuieiras reserve's area (in total, 22735
178 ha) (Zanchi et al., 2014). The valley soils (podzols, gleysols) differ markedly from the clayey
179 plateau and slope soils (oxisols, ultisols). Bleached quartz sand and high amounts of phenolics
180 characterize the valley's podzols (Monteiro et al., 2014; Zanchi et al., 2014; Marques et al.,
181 2016), and the presence of *Mauritia flexuosa* (Luizão et al., 2004), a palm species indicative of
182 hydromorphic conditions, suggests poor drainage (Junk, 1993). A second-order blackwater
183 stream (Rio Açu) drains the area, meeting with Rio Cuieiras, Rio Branquinho, and Rio Negro
184 downstream (Monteiro et al., 2014). The catchment has an area of 660 ha (Monteiro et al., 2014).

185 Monteiro and coworkers report a total annual rainfall of 2806 mm in 2002 and 2004 mm in 2003.
186 Stream discharge was 0.08 – 5.59 m³ s⁻¹ (average, 0.18 m³ s⁻¹) and DOC levels ranged from
187 3.2 – 15.2 mg C l⁻¹ (average, 8.7 ± 3.0 mg C l⁻¹) during that period. The watershed's annual
188 total carbon flux was estimated as 13.3 g C m⁻² in 2003 (Waterloo et al., 2006) and Monteiro
189 and coworkers estimated a stream DOC flux (Rio Açu) of 8.7 g C m⁻² for the same year. Zanchi
190 and coworkers estimated the annual DOC export of the Açu watershed to lie within the range 9.3
191 – 22.7 g C m⁻² (Zanchi et al., 2015). Water samples were taken from piezometers installed
192 across a valley transect (Monteiro et al., 2014) which is maintained and sampled regularly.
193 Piezometers were emptied once before final sampling. The stream was sampled manually, with
194 nitrile gloves, against the direction of flow, using pre-cleaned (acidified ultrapure water, pH2,
195 HCl, Merck EMSURE®, p.a., ACS grade) Nalgene™ polycarbonate bottles (Fisher Scientific,
196 Schwerte, Germany) that were cleaned with the respective sample before final sampling. We also
197 sampled two deep wells on the plateau (35 m and 39 m depth) by lowering an empty, clean
198 sampling bottle on a string until water was reached.

199 The Reserva Campina is a 900 ha reserve that shows only small relief; poor sandy soils (up to
200 99% sand) that co-occur with typical but specific forest types, so-called Campina and
201 Campinarana forests (heath forests; Demarchi et al., 2018) characterize the area. In contrast to
202 the highly diverse plateau (terra firme) forests, Campinas show much lower species diversity and
203 low canopy (~10 m vs. 25-40 m at Reserva Cuieiras). Bare patches of sand cover ~11% of the
204 area (Zanchi et al., 2014). The headwater area is drained by a single headwater stream that is less
205 than 1m wide and often less than 30 cm deep (McClain et al., 1997) and feeds the Rio Tãruma
206 Açu in the southward direction, meeting with the Rio Negro close by Manaus (Figure 1). The
207 catchment area is estimated by 6.5 ha and thus ~100 times smaller than the Rio Açu catchment

208 (Zanchi et al., 2015). According to the same authors, annual rainfall levels are comparable
209 among both reserves. Surface runoff (from the stream) and estimated groundwater outflow
210 amounted to 485 and 1071 mm in that period, respectively (in sum 1556 mm; as compared to
211 Açu 1362 mm; Waterloo et al., 2006). Zanchi and coworkers estimated the total watershed's
212 DOC export with 49.2 g C m⁻², including groundwater outflow and rainfall, of which the stream
213 exported 15.3 g C m⁻². An annual streamflow of 485 mm equals an estimated discharge of 1.0 x
214 10⁻³ m³ s⁻¹, which is about 180 times lower than average discharge at Rio Açu in 2002 – 2003
215 (Monteiro et al., 2014). McClain and coworkers reported annual DOC exports of 40 g C m⁻² yr
216 (McClain et al., 1997; Zanchi et al., 2015) from the same catchment in the period 1993 – 1994.
217 We took samples at the side slopes of the stream from piezometers installed in 1993, as detailed
218 in McClain et al. (1997). For this purpose, wells were emptied three times and sampled afterward
219 (Zanchi et al., 2015). The stream was sampled as described above.

220 2.2 *Water chemistry: TOC, pH, electrical conductivity*

221 Aliquots of the samples were subjected to TOC analysis in the water laboratory of the
222 Instituto Nacional de Pesquisas da Amazônia (INPA) in Manaus, Brazil (Laboratório de Águas
223 do INPA/ CPRHC – Coordenação de Pesquisas em Recursos Hídricos e Clima). Samples were
224 measured on a total organic carbon analyzer (TOC-VCPH model, Shimadzu, Kyoto, Japan)
225 (Monteiro et al., 2014). Before extraction, we analyzed samples for pH and electrical
226 conductivity (EC) with a Multi 340i probe system (WTW, Weilheim, Germany).

227 2.3 *Solid-phase extraction of DOM*

228 DOM samples were solid-phase extracted (SPE) shortly after sampling at INPA, Manaus
229 (Laboratório de Ecossistemas Aquáticos) using an established protocol (Dittmar et al., 2008).
230 The solid-phase sorbent was a modified styrene-divinylbenzene polymer (PPL Bond Elut™,

231 Agilent, Santa Clara, CA, USA). Samples were acidified to pH 2 with 37% hydrochloric acid
232 (Merck EMSURE®, p.a., ACS grade) before extraction. Solvents used for extraction were
233 ultrapure water, acidified ultrapure water (pH 2, HCl), and ultrapure methanol (Biotec Reagentes
234 Analíticos, p.a., ACS grade). We loaded columns with maximal amounts of 3 mg C. The
235 extraction efficiency (EE) of samples with high DOC concentrations ($> 2 \text{ mg l}^{-1}$, $n = 11$) was
236 always $> 60\%$ and on average $71 \pm 8\%$, at a loading ratio of 443 ± 143 (average \pm standard
237 deviation; PPL: DOC in mg/ mg). Sample PT6 was an exception (high DOC, low EE: 33%).
238 Samples with low DOC concentrations ($< 2 \text{ mg l}^{-1}$, $n = 4$) showed lower extraction efficiency
239 ($23 \pm 15\%$), and loading ratios were generally higher (2400 ± 960) (Supporting Information
240 Table S1).

241 2.4 Water isotopes

242 Isotopic signatures of water ($\delta^{2}\text{H}$ - and $\delta^{18}\text{O}$ -values) were analyzed by high-temperature
243 conversion-isotope ratio mass spectrometry (HTC-IRMS) in the stable isotope laboratory of the
244 Max Planck Institute for Biogeochemistry (BGC-IsoLab). For method details, we refer the reader
245 to the literature (Gehre et al., 2004). In short, measurements were conducted on a Delta+ XL
246 coupled to a high-temperature furnace via a ConFlow III interface (Thermo Fisher Scientific,
247 Bremen, Germany). One μl of water was injected using an A200S autosampler (CTC Analytics
248 AG, Zwingen, Switzerland), and the furnace temperature was held at $1350 \text{ }^{\circ}\text{C}$. The $\delta^{2}\text{H}$ and
249 $\delta^{18}\text{O}$ values are reported on the VSMOW scale, which is realized by parallel analysis of samples
250 against in-house standards. In-house standards are routinely calibrated against internationally
251 accepted water standards. Daily standard deviations for $\delta^{2}\text{H}$ and $\delta^{18}\text{O}$ measurements are usually
252 better than 1 and 0.1 ‰, respectively. We accessed regional isotope ratio estimates of rain for
253 October and November with the help of the online isotopes in precipitation calculator (OIPC2.2,

254 version 3.1, <http://wateriso.utah.edu/waterisotopes/>; Bowen and Revenaugh, 2003). The values
255 were similar for both sites; -7 ‰ and -17 ‰ (V-SMOW) for $\delta^2\text{H}$, and -2.4‰ and -3.8‰ (V-
256 SMOW) for $\delta^{18}\text{O}$ in October and November, respectively. We estimated the average between
257 both monthly values for each isotope, considering our sampling date at the end of October/
258 beginning of November.

259 2.5 Radiocarbon analysis of solid-phase extracted DOC

260 We transferred an aliquot of methanolic PPL extract equivalent to 0.25 mg C to tin capsules
261 (8 mm diameter, 20 mm height; IVA Analysentechnik, Meerbusch, Germany). The methanol
262 was left to evaporate. Air-dried capsules were combusted in an elemental analyzer and
263 graphitized for radiocarbon analysis on a 3 MV Tandemron 14C-AMS (HVVEE, Amersfoort,
264 Netherlands) at the Max Planck Institute for Biogeochemistry in Jena, Germany. Modern (Oxalic
265 Acid II) and 14C-depleted standard materials were carried along for quality control and data
266 corrections (Steinhof et al., 2017; Benk et al., 2018). Graphitization is conducted with Duran
267 glass tubes at a temperature of 550°C in the presence of hydrogen and with an iron catalyst
268 (Steinhof et al., 2017). Radiocarbon concentrations are given as fraction modern (F14C), which
269 is the fraction of the standard concentration normalized for $\delta^{13}\text{C}$ of the oxalic acid standard
270 measured at the same time (with $\delta^{13}\text{C}$ of -19‰; Trumbore et al., 2016). $\Delta^{14}\text{C}$ values also take
271 into account radioactive decay of the oxalic acid standard since 1950, which demarks the start of
272 bomb tests that increased the 14C content of the atmosphere (Trumbore et al., 2016). We
273 calibrated our background-corrected F14C data with atmospheric data published in Graven
274 (2015) to derive DOC age. We used values for the inner-tropical convergence zone (zone SH3).
275 To cover the time point of sampling (2017), we predicted $\Delta^{14}\text{C}$ values for 2016 and 2017 based
276 on data from 1997-2015 by linear regression ($R^2 = 0.996$) and then estimated DOC mean

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

279 *2.6 Orbitrap measurements of DOM and data processing*

280 We conducted Orbitrap measurements at the Max Planck Institute for Biogeochemistry, as
281 described elsewhere (Simon et al., 2018), at a nominal resolution setting of 480,000. Allowable
282 numbers of atoms in formula calculation were as follows: 12C, 0-60; 1H, 0-120; 14N, 0-2; 32S,
283 0-1; 16O, 1-60, 13C, 0-1. The assignment was done at ± 1 ppm tolerance. We defined noise as the
284 smallest peak in the sample set, not including zeros. We then calculated the maximal signal-to-
285 noise ratio (S/N_{max}) of a peak across all samples (without blanks and reference material) and
286 only kept peaks with a value > 5 . We calculated the number of matches of each peak across all
287 samples (without blanks and reference) and only kept peaks with more than one match. To
288 exclude contaminants, we calculated a signal-to-noise ratio of blank peaks (only blank samples)
289 and discarded all peaks with values > 20 . Finally, we excluded those peaks that were only
290 present in less than 20% of all samples (i.e., < 9 matches) and small (SN_{max} < 20). After the
291 exclusion of peaks outside the usual mass defect range of natural organic compounds (nominal
292 m/z in Da, -0.05 mDa ... +0.3 mDa), 13658 peaks remained in the dataset in total. Of those,
293 9893 had a formula assigned, and 3672 included one 13C atom. Formulae having H/C ratios > 2
294 or DBE-O values (double bond equivalent minus oxygen atoms) > 15 (modulus) or containing
295 the elemental combination N₂S were discarded due to their unlikely occurrence in DOM
296 (Hawkes et al., 2016). In the case of ambiguous peaks with more than one assigned molecular
297 formula, only suggested CHO formulae with a DBE-O of < 10 (modulus) were kept (Herzprung
298 et al., 2014). Other formula suggestions or ambiguous hits were excluded from the formula pool
299 and kept as “no reference” peaks. 13C-containing formulae that were missing their equivalent

300 monoisotopic (only ^{12}C) formula were excluded. As a last measure, we only considered peaks
301 detected twice in two separate runs for further analysis (Riedel and Dittmar, 2014). The final
302 dataset contained 7705 formulae (of those, 1963 containing a ^{13}C). For comparison of samples,
303 we normalized all mass spectra to the sum of their peak intensities (including all peaks $> \text{S/N} =$
304 5, also those with no assigned formula). The further analysis of the data focused on the subset of
305 peaks with an assigned monoisotopic formula ($n = 5709$). The crosstab is available from
306 <https://doi.org/10.1594/PANGAEA.922606> (Supporting Information Data Set S1).

307 *2.7 Statistical analyses: Ecosystem fingerprints and molecular links*

308 We analyzed the molecular formula data by Principal Coordinate Analysis (PCoA, cmdscale
309 function, stats package, v3.5.1) and post-ordination gradient fitting analysis (envfit function,
310 vegan package, v2.5-2) within the statistical computation environment R Studio (v1.1.453, ©
311 2009-2018 RStudio, Inc.). PCoA was based on Bray-Curtis dissimilarities obtained by the
312 function vegdist from package vegan (Oksanen, 2010). The method allows the comparison of
313 samples based on their formula populations. Redundancy within the dataset, i.e., formulae
314 showing similar trends in ion abundance across samples, is effectively reduced and yields a set of
315 coordinates that summarize the variability of the dataset best (Osterholz et al., 2016). We then
316 analyzed the distribution of samples in coordinate space for clustering (indicating similarity
317 among samples) and correlations with specific DOM indices by the envfit function (at 999
318 permutations). DOM indices aggregate properties of the molecular formula population of each
319 sample (Supporting Information Table S2). We further assessed molecular formulae with a
320 significant ($p = 0.05$) association to clusters of samples derived from PCoA analyses by
321 Student's t-test of averaged relative ion abundances across samples of each group. We conducted
322 two-sided tests assuming unequal variances.

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

339 **3 Results**

340 *3.1 Water chemistry, water isotopic composition and carbon isotopic composition of DOC*

341 Samples were clearly differentiated by water chemistry, revealing two major endmembers
342 related to fine- and coarse-textured soils ("Plateau" vs. "Upland"; Figure 2, Supporting
343 Information Table S1), and one intermediate site ("Valley"). Acidity (pH), electrical conductivity
344 (EC) and concentrations of dissolved organic carbon (DOC) were 3.7 ± 0.1 , $49 \pm 6 \mu\text{S cm}^{-1}$ and
345 $37.7 \pm 7.54 \text{ mg C l}^{-1}$ 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

369 samples indicated variation in isotopic composition within the range of the other samples. While
370 the $\delta^{13}\text{C}$ composition of all DOC extracts was roughly similar (average \pm standard deviation: -
371 29.40 ± 0.45 ‰, $n=13$) and reflected a typical C3 source, their calibrated radiocarbon ages were
372 young (less than 2 to 13 years old at maximum) and differed markedly by five years between
373 both sites, with the upland site showing older DOC (Figure 2e).

374 3.2 DOM characterization by ultrahigh-resolution mass spectrometry

375 The analysis of molecular DOM composition reflected the separation of samples based on
376 water and DOC properties presented in the previous section (Figure 3a). The PCoA separated
377 samples into two main clusters (“plateau” and “WSE”) on the first coordinate, which held 78%
378 of molecular variation. Plateau-derived DOM was thus most dissimilar from WSE-DOM, in line
379 with a major control of DOC properties and abundance by soil texture (Remington et al., 2007).
380 Consequently, PCoA 1 was linked to significant (Pearson’ r , $p < 0.05$) trends in pH, EC, and
381 DOC (not shown). In line with more subtle differences between WSE samples, the overall
382 explained variability of the second coordinate was smaller (10%). However, WSE sites were
383 clearly separated, and molecular trends thus paralleled differences in water isotopes and
384 radiocarbon age. Moreover, the separation of WSE-DOM by sites suggested differing trajectories
385 of DOM processing that seemed to converge to a common DOM composition (note the close
386 proximity of samples P7 and PR7 in Figure 3a). Several significant trends in molecular indices
387 derived from the DOM data paralleled the PCoA separation (Figure 3b-d). The used descriptors
388 – a-priori defined molecular groups and formula classes that aggregate molecular composition
389 information – are described in Supporting Information Table S2, and data is available online
390 from <https://doi.org/10.1594/PANGAEA.922606> (Supporting Information Data Set S1). WSE-
391 DOM was more oxidized (higher O/C and NO₃C), less saturated (lower H/C, higher DBE) and

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

408 As expected from the results of the gradient analysis, we found clear sets of markers for each
409 DOM type that paralleled trends of weight-averaged molecular indices (Figure 3b-d). To this
410 end, we separated unique from shared (“common”) markers, and thus excluded non-informative
411 formulae (Figure 4, Supporting Information Figure S4, and Supporting Information Figure S5). It
412 is important to note that our definition of “unique” and “common” relates to significant
413 differences in abundance. In fact, most molecular formulae were shared based only on presence
414 (30% of all formulae shared among all three ecosystems, and 83% among WSEs; Supporting

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

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

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

446 3.3 *Molecular links between terrestrial ecosystems and the aquatic continuum*

447 We compared the sets of distinct DOM markers of the valley and upland WSEs to known Rio
448 Negro markers in order to gain qualitative insight into the contribution of headwater aquifer and
449 stream DOM to the “integrated” watershed signal of the lower Rio Negro (Figure 5). Although
450 many CHNO and CHOS formulae were part of the specific sets of signals in both WSEs, the
451 overlap to Rio Negro markers was restricted to CHO formulae. The two WSEs differed in terms
452 of potential molecular links. The cluster of heavy, highly oxidized, and aromatic upland WSE
453 markers indicated consistent overlap with known Rio Negro markers (Figure 5). Overlapping
454 sets of formulae appeared in a confined area of the Van Krevelen plot (boxes in Figure 5b, d;
455 ranges: H/C 0.4 – 0.8, O/C 0.4 – 0.8), suggesting robust matching. However, the overlap was
456 caused by different sets of Rio Negro markers, as indicated by the formula’s m/z (Figure 5c, f).
457 One study (Gonsior et al., 2016) revealed molecular links with markers of higher mass (m/z 350
458 – 650, Figure 5c), while another (Simon et al., 2019) found them in the lower mass range (m/z
459 200 – 500, Figure 5f). Nevertheless, all these markers were specific to the upland WSE and

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

462 **4 Discussion**

463 *4.1 Site characterization: Biogeochemistry of water and carbon*

464 The general analysis of water chemistry showed clear differences among samples, revealing
465 two major endmembers and one intermediate environment (Figure 2). Strongest differences were
466 apparent between fine- and coarse-textured soils (“Plateau” vs. “Upland”; Figure 2, Supporting
467 Information Table S1). Valley samples PR10, PT06, and stream RA indicated “intermediate”
468 levels in terms of pH, electrical conductivity and DOC concentration and were likely plateau-
469 influenced, i.e., affected by mixing due to plateau water supply, or dilution from precipitation in
470 case of stream RA at the time point of sampling. In contrast, similarly high DOC concentrations
471 of all samples from Reserva Campina draining the upland site indicated direct contact between
472 stream and surrounding soils at the time point of sampling (McClain et al., 1997; Zanchi et al.,
473 2015). Local differences in ecohydrology can explain the heavier isotopic composition of
474 groundwater and stream water in the drier upland Reserva Campina (Supporting Information
475 Figure S2; Zanchi et al., 2014; Zanchi et al., 2015), and the lighter isotopic composition of water
476 in the valley and its adjacent plateaus (Leopoldo et al., 1982; Kunert et al., 2017). While
477 isotopically heavy samples like upland P2 (most upslope) and valley PR8/ PR9 (high water
478 levels) likely reflected the imprint of recent precipitation, groundwater replenished during
479 previous wet seasons can explain the lighter water isotopic composition (Leopoldo et al., 1982;
480 Tomasella et al., 2007; Zhang et al., 2009; Miguez-Macho and Fan, 2012; Jasechko and Taylor,
481 2015; Zanchi et al., 2015; Supporting Information Figure S3). Radiocarbon content differed
482 among sites but indicated very young DOC age. Young radiocarbon ages of tropical DOC are

483 frequently reported from aquatic (Mayorga et al., 2005; Moyer et al., 2013; Ward et al., 2013)
484 and soil systems (James et al., 2019) and are explained by strong linkages between C fixation,
485 DOC release, and nutrient recycling (Mayorga et al., 2005). The slight but consistent differences
486 in DOC radiocarbon ages between the drier upland and the wetter valley system likely relate to
487 processes that govern the short-term (years to decades) turnover of organic matter, for example
488 litter turnover and soil respiration rates (Zanchi et al., 2011; Zanchi et al., 2014; Zanchi et al.,
489 2015).

490 *4.2 DOM properties and fingerprints reflect biogeochemical differences among ecosystems*

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

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

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

529 contained unique, highly oxidized, aromatic, and phenolic CHO formulae (Figure 4g-i). These
530 molecules likely represent the initial decomposition products of plant material containing large
531 portions of lignin, cellulose, tannin, flavonoids, and terpenoids. Under non-stagnating, well-
532 drained conditions, sandy soils favor the escape of such surface signals to streams because of the
533 low sorption capacity of soil (Remington et al., 2007). However, fast export contrasts with older
534 calibrated radiocarbon ages of DOC at the upland WSE (~11 years) as compared to the valley
535 WSE site. This delay can be explained by slower litter turnover and reduced CO₂ efflux rates as
536 discussed above (Zanchi et al., 2011; Zanchi et al., 2014). Highly acidic conditions, periods of
537 drought, and an adapted plant community emerge as main drivers of such ecosystem-level
538 differences. It is of special interest that fungi are known to remain active under dry and acidic
539 conditions (Rousk et al., 2010; Vasco-Palacios et al., 2018). Fungal enzymes can potentially alter
540 DOM towards higher-molecular weight, aromatic, and oxidized structures (Waggoner et al.,
541 2015; Zavarzina et al., 2018), all of which seem to be occurring uniquely in upland WSE soils.

542 *4.3 Upland Campina WSEs are a potential source of indicative Rio Negro markers*

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

552 can easily affect the instrument response in terms of ion abundance patterns and thus, m/z range
553 (Hawkes et al., 2016; Simon et al., 2018; Hawkes et al., 2020). Despite sample set and lab/
554 instrument effects, measurement settings were relatively similar among the two available studies
555 and ours, thus supporting the general overlap of the upland WSE marker and Rio Negro marker
556 sets in chemical space (oxidation and saturation state; Supporting Information Text S1).
557 Accumulation time of ions before FT-MS analysis is one factor that explains m/z shifts (Hawkes
558 et al., 2016; Simon et al., 2018). However, sampling criteria such as location, season, scale, and
559 coverage will also have a strong influence on the resulting sets of markers. While it is not
560 surprising to see large differences in marker sets (Supporting Information Figure S8 and
561 Supporting Information Figure S9), it is remarkable to find small but consistent overlap between
562 studies, pointing toward subsets of markers with specific oxidation and saturation states. It is
563 known that different instruments capture gradients of variation across a sample set even at very
564 slight compositional differences (Hawkes et al., 2016), as may be expected when samples
565 originate from similar environmental contexts as in our study. It is thus highly encouraging to
566 find consistent and robust molecular overlap in three independent FT-MS datasets, being in line
567 with long-standing hypotheses of landscape functioning in the Rio Negro basin (Leenheer, 1980;
568 Goulding et al., 1988), namely that whitesand ecosystems are potential sources of indicative Rio
569 Negro markers. Counter-intuitively, the highly indicative set of upland WSE markers were not
570 only found to overlap with Rio Negro samples from the proximity of its draining higher-order
571 river, the Rio Tarumã Açu (Simon et al., 2019) but also in samples upstream (Gonsior et al.,
572 2016; compare Figure 1). This observation implies that similar high-molecular-weight oxidized
573 aromatic compounds are exported upstream of Novo Airão, possibly by other upland, Campina-

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

576 To identify potentially robust candidate markers, we queried the nine upland whitesand Rio
577 Negro markers that consistently matched with the two independent lists of Rio Negro markers
578 (“Upland overlap”, red squares in Figure 5) in PubChem to obtain potential structure suggestions
579 (Supporting Information Table S3). A query with the ChEBI (chemical entities of biological
580 interest) database yielded only two suggestions that were also found in PubChem. The structural
581 suggestions thus represent only a first indication of potential structures and not necessarily
582 structures that were identified in DOM or organisms. Pubchem yielded up to 108 hits for formula
583 $C_{10}H_6O_6$ and five at minimum for formula $C_{11}H_6O_8$ (numbers after exclusion of twelve ions, one
584 hydrate and six stereoisomers from the list). In general, structures were highly oxidized, showing
585 on average 2.5 carbonyl groups, 0.97 ether bonds, and 0.65 lactone groups per molecule across
586 formulae. All formulae were classified as Black-Carbon like or polyphenol-like (BC, PP). On
587 average, 68% of suggested structures per formula featured at least one heterocyclic ring, with
588 five and six-membered rings dominating (each 30% on average). Functional group count was
589 strongly correlated with mass ($R^2 = 0.84$, range = 5.2 – 9.7), same as numbers of carbon double
590 bonds (0.86, 3.2 – 8.4) aromatic rings per molecule (0.87, 0.6 – 2.6) and thus, DBE (0.85, 8 –
591 16). Suggestions of smaller molecules (< 270 Da, with less than 15 C atoms) were by tendency
592 dominated by naphthalene and chromene structures that are oxygen-poor scaffolds consisting of
593 two rings. Larger molecules > 320 Da (with >15 C atoms) in contrast showed more hits
594 involving at least one benzofuran, phenol or benzenecarboxylic scaffold, which all show a higher
595 degree of oxidation and functionalization. Chromenes and benzofurans are well known photo-
596 and bioactive plant metabolites (Towers and Hudson, 1987). Knowledge of these trends and

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

602 Groundwater DOM markers revealed a potential direct link between ecosystem-specific
603 headwater and ecosystem-integrated downstream signals in the Rio Negro basin. The consistent
604 overlap of upland WSE and Rio Negro markers was unexpected because our initial hypothesis,
605 namely that typical valley WSE's will show stronger, or at least similar, matching with Rio
606 Negro markers, had to be refused. Highly specific, tannin-like aromatic DOM compounds of the
607 smaller upland WSE watershed represent a potential molecular link between headwaters and the
608 Rio Negro. This finding opposes the assumed role of riparian valley WSEs as main sources of
609 DOM in the Rio Negro watershed (McClain et al., 1997; Bardy et al., 2011). Although Bardy
610 and coworkers showed that upland Campina forests areas produce a marked DOM signal
611 traceable to rivers, others found that riparian and wetland WSEs along rivers contributed mainly
612 to DOC exports of Amazonian and other watersheds (Dosskey and Bertsch, 1994; McClain et al.,
613 1997; Remington et al., 2007). Our results suggest that there is more detail to add to this simple
614 model. Small watersheds with strong terrestrial-lotic linkages may leave a more significant
615 downstream imprint as expected from absolute annual discharge, areal extent, or DOC export.
616 Our results show that highly specific forest ecosystems potentially leave a distinct imprint within
617 the Rio Negro's exported DOM, although contributing only secondarily to the overall fluxes of
618 water, and maybe, DOC. The smaller upland WSE watershed at Reserva Campina shows a two
619 times lower average annual stream discharge (based on the catchment area) but a two-three times

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

637 **5 Conclusion**

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

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

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

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

677 **Acknowledgments**

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

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

691 **Declaration of potential conflicts of interest**

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

694 **Data availability**

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

697 **Supplementary Material**

698 29 pages, containing 9 Tables, 17 Figures and 17 References.

699 **References**

- 700 Abbott B. W., Gruau G., Zarnetske J. P., Moatar F., Barbe L., Thomas Z., Fovet O., Kolbe T.,
701 Gu S., Pierson-Wickmann A. C., Davy P. and Pinay G. (2018) Unexpected spatial stability
702 of water chemistry in headwater stream networks. *Ecol. Lett.* **21**, 296–308.
- 703 Adeney J. M., Christensen N. L., Vicentini A. and Cohn-Haft M. (2016) White-sand ecosystems
704 in Amazonia. *Biotropica* **48**, 7–23.
- 705 Alvarez-Cobelas M., Angeler D. G., Sánchez-Carrillo S. and Almendros G. (2012) A worldwide
706 view of organic carbon export from catchments. *Biogeochemistry* **107**, 275–293.
- 707 Bardy M., Derenne S., Allard T., Benedetti M. F. and Fritsch E. (2011) Podzolisation and
708 exportation of organic matter in black waters of the Rio Negro (upper Amazon basin,
709 Brazil). *Biogeochemistry* **106**, 71–88.
- 710 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

757 2H/1H and 18O/ 16O analysis of water samples with dual inlet precision. *Rapid Commun.*
758 *Mass Spectrom.* **18**, 2650–2660.

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

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

766 Graven H. D. (2015) Impact of fossil fuel emissions on atmospheric radiocarbon and various
767 applications of radiocarbon over this century. *Proc. Natl. Acad. Sci.* **112**, 9542–9545.

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

770 Hawkes J. A., D’Andrilli J., Agar J. N., Barrow M. P., Berg S. M., Catalán N., Chen H., Chu R.
771 K., Cole R. B., Dittmar T., Gavard R., Gleixner G., Hatcher P. G., He C., Hess N. J.,
772 Hutchins R. H. S., Ijaz A., Jones H. E., Kew W., Khaksari M., Lozano D. C. P., Lv J.,
773 Mazzoleni L., Noriega-Ortega B., Osterholz H., Radoman N., Remucal C. K., Schmitt N.
774 D., Schum S., Shi Q., Simon C., Singer G., Sleighter R. S., Stubbins A., Thomas M. J.,
775 Tolic N., Zhang S., Zito P. and Podgorski D. C. (2020) An international laboratory
776 comparison of dissolved organic matter composition by high resolution mass spectrometry:
777 Are we getting the same answer? *Limnol. Oceanogr. Methods* **accepted**, doi:
778 10.1002/lom3.10364.

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

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

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

785 Herzsprung P., Hertkorn N., von Tümpling W., Harir M., Friese K. and Schmitt-Kopplin P.
786 (2014) Understanding molecular formula assignment of Fourier transform ion cyclotron
787 resonance mass spectrometry data of natural organic matter from a chemical point of view.
788 *Anal. Bioanal. Chem.* **406**, 7977–7987.

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

791 Hutchins R. H. S., Aukes P., Schiff S. L., Dittmar T., Prairie Y. T. and del Giorgio P. A. (2017)
792 The Optical, Chemical, and Molecular Dissolved Organic Matter Succession Along a
793 Boreal Soil-Stream-River Continuum. *J. Geophys. Res. Biogeosciences* **122**, 2892–2908.

794 James J. N., Gross C. D., Dwivedi P., Myers T., Santos F., Bernardi R., Fidalgo de Faria M.,
795 Amaral Guerrini I., Harrison R. and Butman D. (2019) Land use change alters the
796 radiocarbon age and composition of soil and water-soluble organic matter in the Brazilian
797 Cerrado. *Geoderma* **345**, 38–50.

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

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

802 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

918 drinking water treatment to its molecular characteristics. *Water Res.* **113**, 149–159.

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

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

928 Remington S. M., Strahm B. D., Neu V., Richey J. E. and Da Cunha H. B. (2007) The role of
929 sorption in control of riverine dissolved organic carbon concentrations by riparian zone soils
930 in the Amazon basin. *Soil Sci.* **172**, 279–291.

931 Riedel T. and Dittmar T. (2014) A Method Detection Limit for the Analysis of Natural Organic
932 Matter via Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.*
933 **86**, 8376–82.

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

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

939 Roth V.-N., Lange M., Simon C., Hertkorn N., Bucher S., Goodall T., Griffiths R. I., Mellado-
940 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., Elozegi 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., Selmeçzy 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

1033 isotopic effect. *Int. J. Climatol.* **29**, 619–628.

1034 **Captions**

1035 Figure 1. Detail of the lower Rio Negro catchment northwest of Manaus showing sampling sites
1036 of available FT-MS studies. We sampled at two locations north of Manaus (green dots;
1037 “Cuieiras”, “Campina”). Two other FT-MS datasets from blackwater sampling locations are
1038 shown, including the Rio Negro and connected lakes (black dots; Gonsior et al., 2016) and the
1039 lower reach of the Rio Negro and two of its tributaries close to Manaus (red dots; Simon et al.,
1040 2019). We accessed watershed limits and river data as shapefiles from www.ore-hybam.org
1041 (Seyler et al., 2009). The headwater stream width is not drawn to scale. Roads and main water
1042 bodies were extracted from the OpenStreetMap project (OSM; natural features and roads) and
1043 downloaded as shapefiles from www.download.geofabrik.de. Map editor: Marcus Guderle, MPI
1044 Jena.

1045 Figure 2. Differences in the water and SPE-DOM properties. Variables shown are a) pH, b)
1046 electrical conductivity, c) dissolved organic carbon concentration, d) difference in $\delta^{18}\text{O}$ values
1047 of water compared to regional average precipitation (OIPC estimate), e) mean calendar age of
1048 DOC in solid-phase extracts. Samples are grouped into biogeochemical environments (upland
1049 Campina; riparian valley, and plateau). Symbols denote sample type (squares, piezometer; star,
1050 deep well; circle, stream); open symbols mark “intermediate” valley samples (see main text).

1051 Figure 3. Multivariate analysis based on molecular DOM data. a) Separation of samples in a
1052 principal coordinate analysis (PCoA) based on Bray Curtis dissimilarity. The plot shows only the
1053 first two coordinates, the third coordinate (5% of explained variability, not shown) did not
1054 contribute further to separation. Percentages denote the degree of explained variability in DOM
1055 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

1079 (common to all samples) are shown in Supporting Information Figure S4. Left column panels (a,
1080 d, g) show the average (and max) mass spectrum of each ecosystem. Mid column panels (b, e, h)
1081 show the formula subsets in Van Krevelen space (each dot represents a molecular formula
1082 defined by its atomic ratios of hydrogen, H/C, and oxygen to carbon, O/C). Formulae are colored
1083 according to classes (see legend and numbers of formulae in panels to the right). The plot is
1084 divided by solid lines that mark molecular group categories (see also caption of Figure 3, and
1085 Supporting Information Table S2). Right column panels (c, f, i) show the same formula subsets
1086 from mid-row panels in H/C vs. m/z (mass-to-charge) space. Dotted grey lines are for visual
1087 guidance and comparison (at $m/z = 450$, $O/C = 0.5$, and $H/C = 1$). Left- and right-column plots
1088 share the m/z abscissa (x-axis) while mid- and right-column plots share the H/C ordinate (y-
1089 axis).

1090 Figure 5. Overlap of WSE-specific formulae (significantly enriched formulae) with two
1091 independent sets of Rio Negro DOM markers (a-c: data from Gonsior et al., 2016; d-f: data from
1092 Simon et al., 2019). Left panels (a, d) show the overlap of three sets of markers (Rio Negro,
1093 Valley, Upland) in a Venn diagram created with package “eulerr” in RStudio. Overlap indicates
1094 common formulae, i.e., shared information. Areas are scaled to number of formulae, which are
1095 given for each subset (note color, area may be dissected). Overlap is shown only for CHO
1096 formulae because Rio Negro markers were mainly CHO-type, other formula classes showed no
1097 distinct overlap. Right panels show Rio Negro markers of each study (grey filled squares) in
1098 chemical space as a Van Krevelen plot (b, e) or H/C vs. m/z plot (c, f; similar visualization as in
1099 Figure 4). Colored symbols show the match between datasets (“overlap”) and refer to the bold
1100 ellipse in panels a and d, respectively. Plateau markers are additionally added for comparison.
1101 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).