Supplementary Material

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

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

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

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

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

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

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

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

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

Contents

Introduction

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

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

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

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

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

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

Supplementary Table S1. Combined data of samples described in this study. All samples were taken in 2017. Greyed entries denote problematic data (see additional comments below table). Abbreviations: EC, electrical conductivity; F14C, Fraction Modern; Δ14C, correction accounting for decay between sample collection and measurement; EE, extraction efficiency based on DOC of samples and SPE extracts.

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

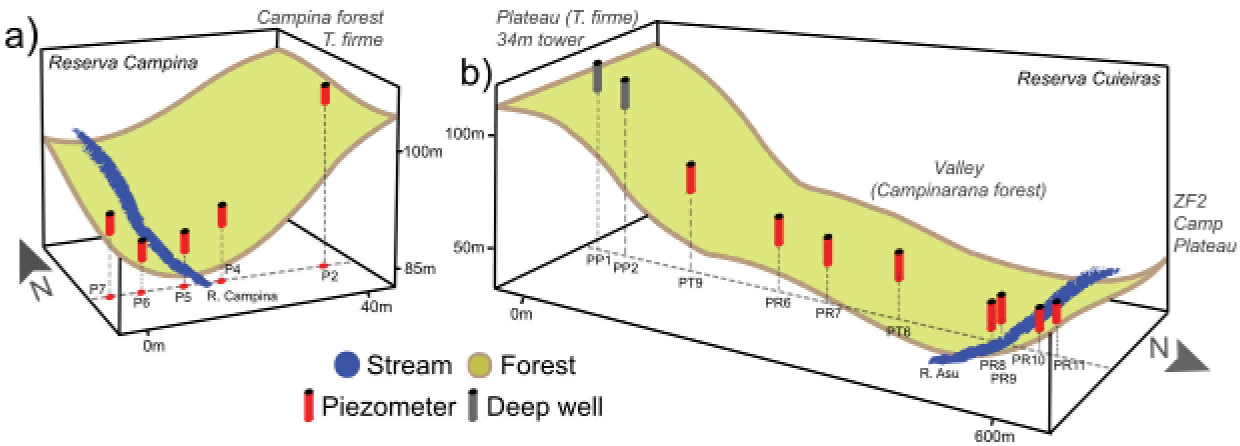
* Reserva Cuieiras: In piezometers, water level below the surface at sampling (daily mean, hourly data), in wells: maximum depth. Reserva Campina: piezometers, max. depth. ** n.d., not determined due to the limited amount of extract. # Value likely influenced by ^{14}C-dead contaminant signal. ## Calculated based on the formula d-excess = δ2H – 8* $\delta^{18}O$ (Dansgaard, 1964).

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

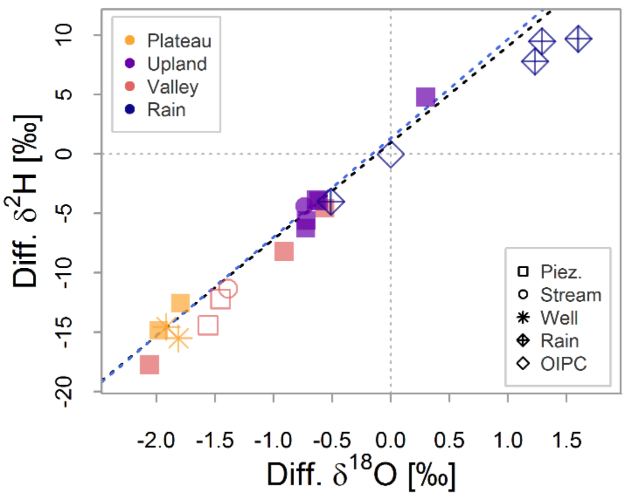
Index	Explanation	Calculation/ definition	Reference of use
H/C	Atomic ratio of hydrogen to carbon in a molecular formula	H/C	Kew et al., 2017; Kim et al., 2003
O/C	Atomic ratio of oxygen to carbon in a molecular formula	O/C	
DBE	Double Bond Equivalents	1+0.5*(2*C-H+N+P)	Koch and Dittmar, 2016, 2006
AImod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

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

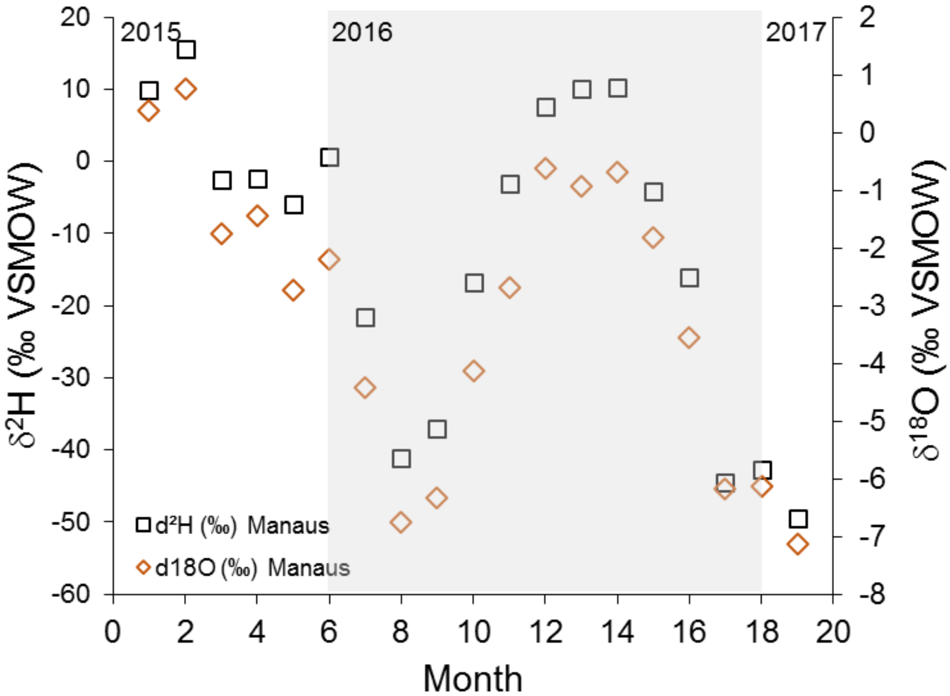
Formula	$C_{10}H_6O_6$	$C_{11}H_6O_6$	$C_{11}H_6O_7$	$C_{11}H_6O_8$	$C_{12}H_{10}O_7$	$C_{17}H_8O_7$	$C_{21}H_{12}O_8$	$C_{22}H_{14}O_9$	$C_{20}H_{14}O_{11}$
Ecosystem	UPL	WSE	UPL	UPL	WSE	UPL	WSE	UPL	UPL
Indices based on molecular formula									
Molecular group	BC	BC	BC	BC	PP	BC	BC	PP	PP
Mass [Da]	221	233	249	264	265	323	391	421	429
DBE	8	9	9	9	8	14	16	16	14
AI_{MOD}	0.71	0.75	0.73	0.71	0.53	0.78	0.71	0.66	0.59
O/C	0.60	0.55	0.64	0.73	0.58	0.41	0.38	0.41	0.55
H/C	0.60	0.55	0.55	0.55	0.83	0.47	0.57	0.64	0.70
Hits (PubChem)	108	40	16	5	96	20	18	18	7
PubChem: Fused rings									
Average per molecule	1.88	2.28	1.69	1.40	1.58	2.45	3.28	2.61	1.71
Hits > one ring [%]	76	88	63	40	52	100	94	56	43
Hits > two rings [%]	11	40	6	0	6	40	61	44	14
PubChem: Aromatic rings									
Average per molecule	0.92	0.95	1.06	0.60	0.96	1.80	2.39	2.56	2.14
Hits > one ring [%]	8	5	6	20	14	80	100	100	100
Hits > two rings [%]	0	0	0	0	0	5	33	56	14
PubChem: Quinone-like rings									
Average per molecule	0.24	0.05	0.19	0	0.18	0.05	0.17	0.06	0
Total hits [%]	17	5	19	0	10	5	11	6	0
PubChem: Oxygen heterocycles									
Average per molecule	0.87	1.35	0.69	1.20	0.64	1.65	2.06	0.89	0.43
Hits > zero rings [%]	68	90	56	80	52	95	89	50	29
Hits > one ring [%]	19	45	13	40	10	70	72	17	14
Hits w pyran-like ring(s) [%]	29	45	25	40	20	15	50	28	0
Hits w furan-like ring(s) [%]	30	50	19	20	18	70	39	17	14
Hits w other O-heterocycle(s) [%]	13	28	19	20	17	20	22	11	29
PubChem: Specific scaffolds/ substructures									
Hits w chromene unit(s) [%]	13	0	19	0	10	0	11	0	0
Hits w naphthalene unit(s) [%]	26	48	19	0	16	10	28	11	0
Hits w benzoic acid/ phenol unit(s) [%]	16	3	31	20	32	0	28	67	100
Hits w benzofuran unit(s) [%]	14	20	13	0	5	70	22	22	29
PubChem: Functional groups (average per molecule)									
Double bonds	3.23	3.55	3.50	3.20	3.61	6.05	8.39	8.39	6.57
Carboxyl (COOH)	0.72	0.80	1.06	1.40	0.97	0.20	0.28	1.61	1.86
Hydroxyl (OH), w/o carboxyl	1.21	0.33	0.81	0.60	1.58	0.60	2.11	1.89	2.00
Carbonyl (C=O), w/o carboxyl	2.02	2.25	2.63	2.60	1.61	3.90	2.56	1.94	3.00
Methyl (Me), w/o methoxy	0.01	0.15	0.06	0	0.25	0	0.06	0.28	0.71
Methoxy (MeO)	0.05	0.08	0.06	0	0.48	0	0.22	0.11	0.29
Lactone (carboxyl ring condensation)	0.57	0.80	0.31	0.80	0.28	1.60	0.67	0.11	0.71
Ether (COC), w/o methoxy & lactone	0.65	0.83	0.81	0.80	0.96	0.40	1.56	1.56	1.14



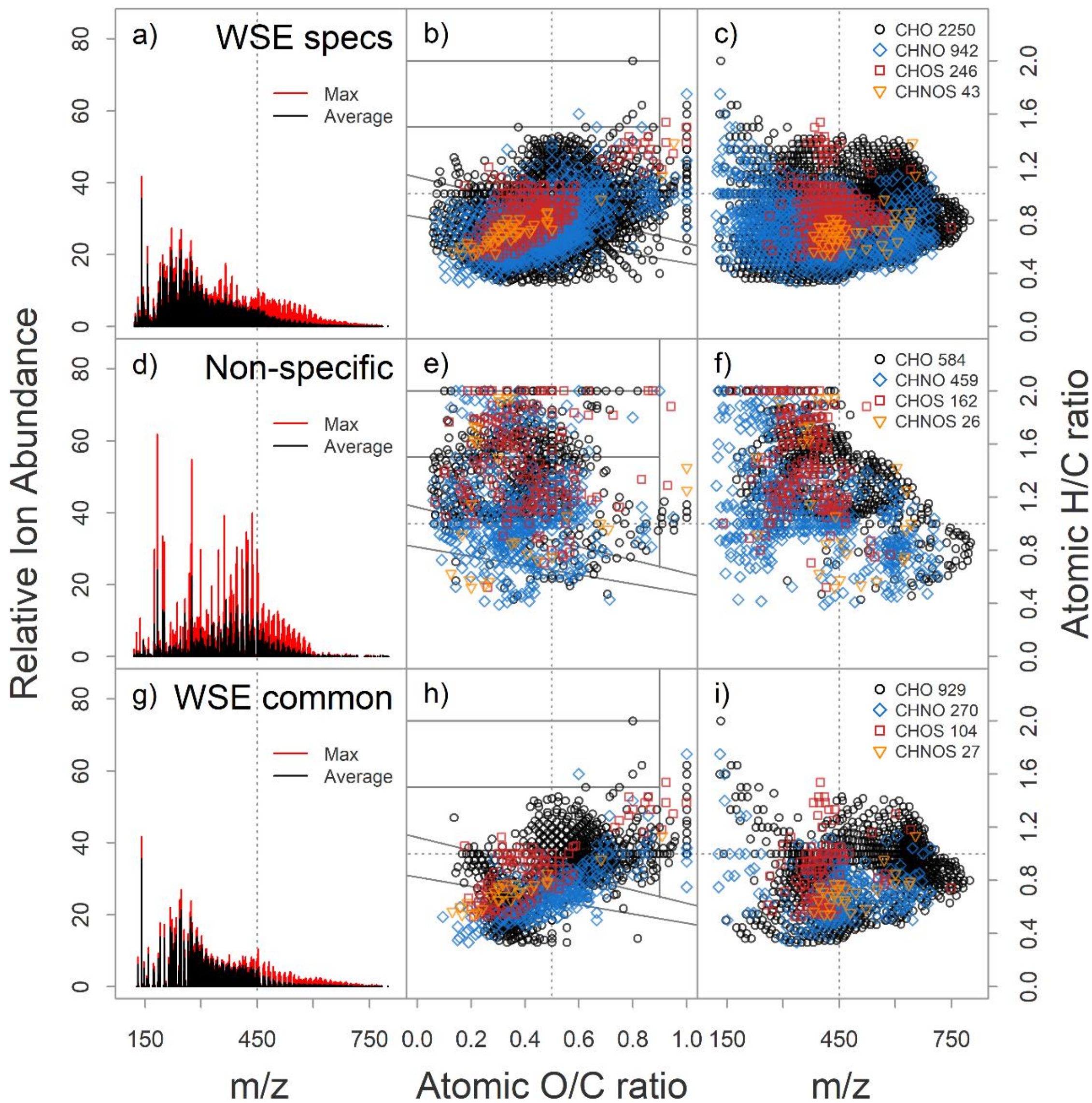
Supplementary Figure S1. Schematic landscape sections of the two sampled whitesand ecosystems with sample locations along transects. a) Upland WSE Campina forest transect at Reserva Campina. b) Elevated plateau with intersected riparian valley WSE at Reserva Cuieiras. Note differences in scale.



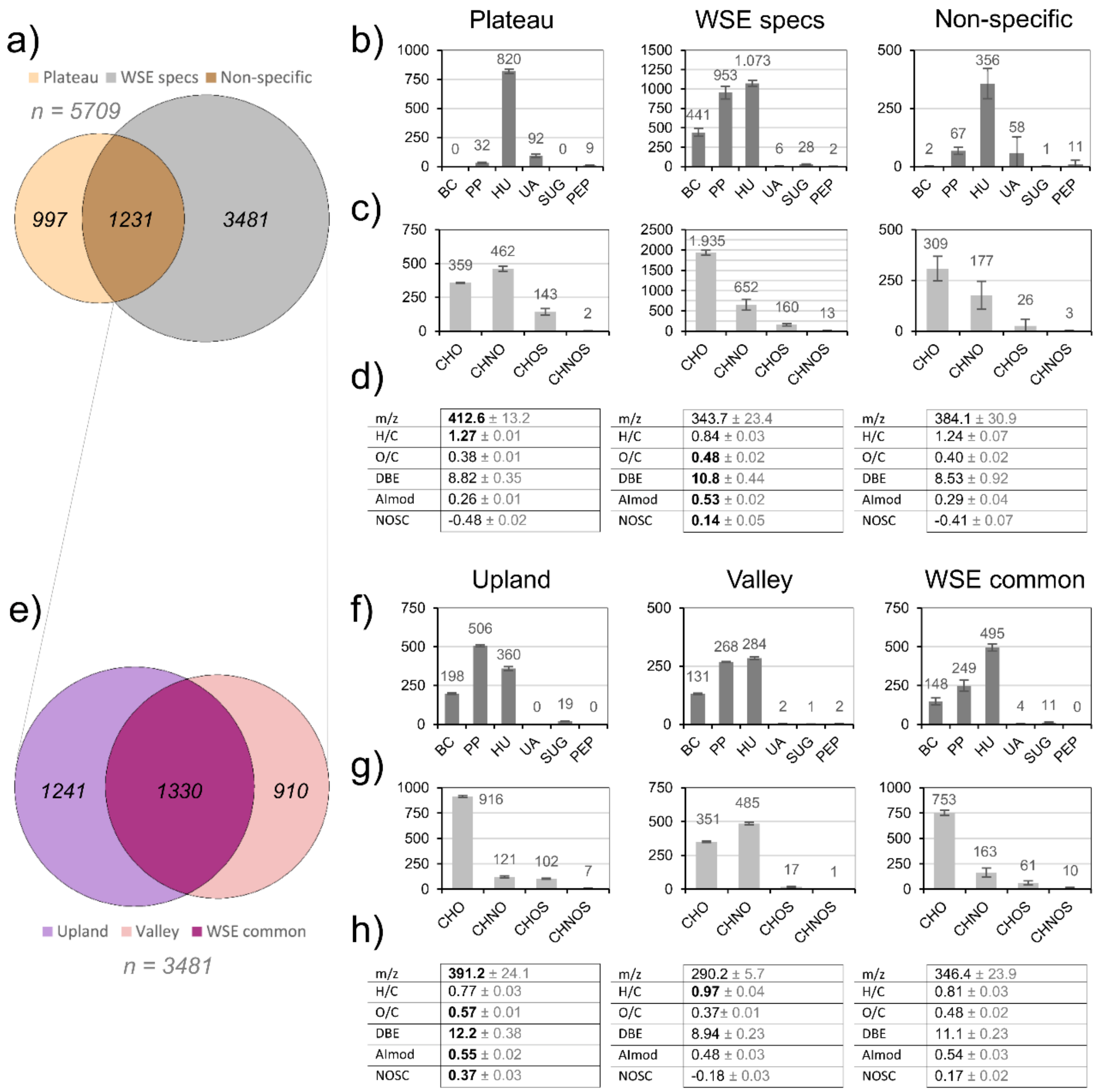
Supplementary Figure S2. Water isotope data in relative delta notation against predicted average precipitation by OIPC (blue open diamond at plot origin; OIPC 3.0 estimate, see methods) in water samples taken in October/ November 2017. Black dotted line: Local meteoric water line (LMWL) constructed from data collected monthly by the INPA climatology station, located in the Adolpho Ducke Forest Reserve, Manaus, Brazil (data 08/15 - 02/17): δ^2H = 8.343*δ^{18}O + 13.362 (r = 0.99, n = 19). Four selected rain datapoints from October and November 2015 and 2016 are shown as blue crossed diamonds. Blue dotted line: LMWL constructed from monthly δ^2H and δ^{18}O in precipitation at IAEA/ WMO (International Atomic Energy Agency/ World Meteorological Organization) station in Manaus, Brazil, by Zhang et al., (2009), data from 1965 – 1990: δ^2H = 8.14*δ^{18}O + 12.96 (r = 0.98, n = 186).



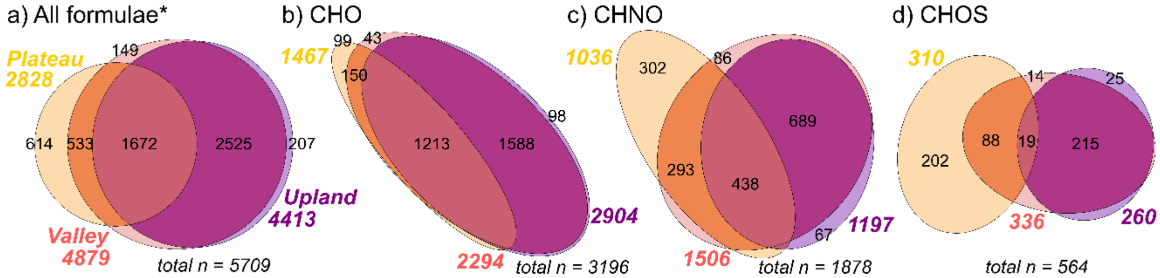
Supplementary Figure S3. Water isotope data from INPA's meteorological station at Reserva Ducke, north of Manaus, over the course of the year (data coverage: 08/15 – 02/17; n=19). Grey shading marks 2016; from 01/2016 – 01/2017. Water is isotopically light during the wet season (February - May), and becomes heavy in the dry season, peaking from July- September.



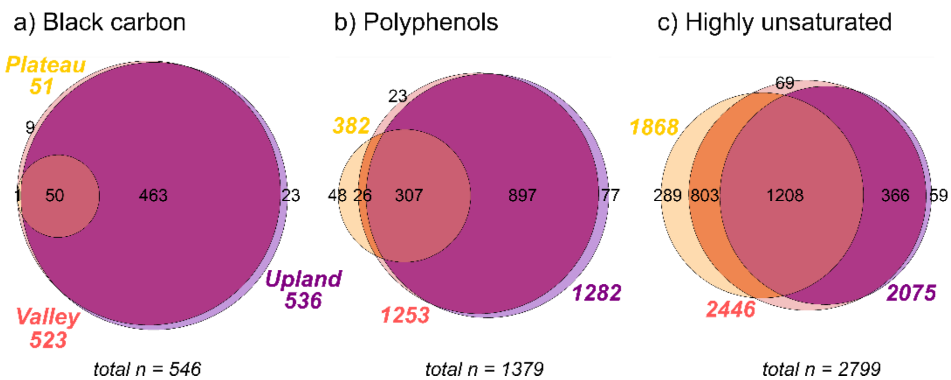
Supplementary Figure S4. Subsets of non-informative molecular formulae (not related uniquely to one of the three biogeochemical environments but common to two or three). Left column panels (a, d, and g) show the average (and max) mass spectrum of each sample set. Mid column panels (b, e, and h) show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



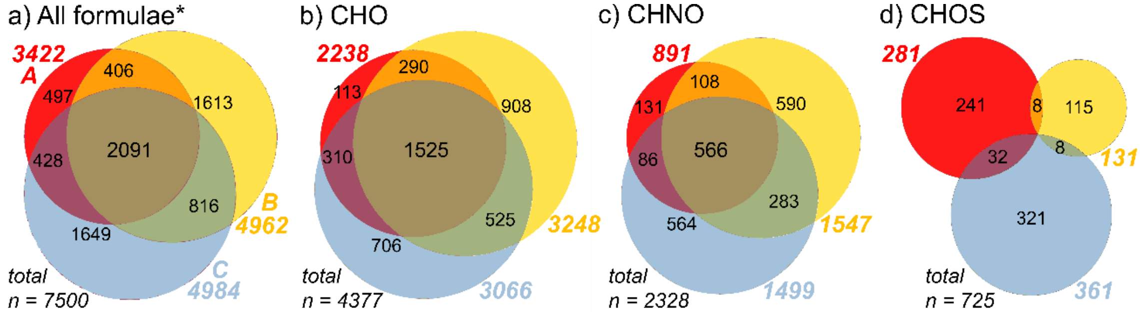
Supplementary Figure S5. Overview of all subsets of molecular formulae, showing results of the comparison among samples from all three biogeochemical environments (a – d) and the WSE sites (Upland and Valley samples) only (d – h). Names of subsets relate to data shown in Figure 4 and Figure S4. Venn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



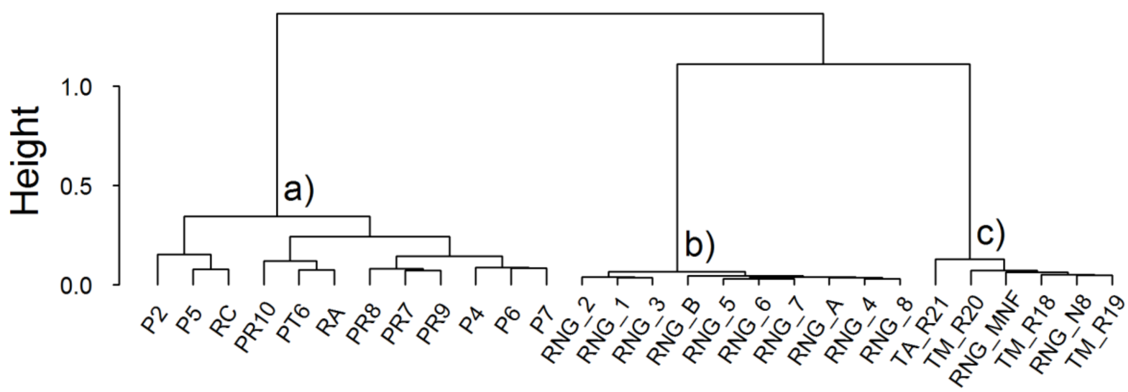
Supplementary Figure S6. Venn diagrams showing the overlap between molecular formulae between subsets of DOM samples, for a) all formulae (*including CHNOS formulae), b) formulae containing only carbon, hydrogen and oxygen atoms, c) formulae containing one or two N atoms, and d) formulae containing an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



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



Supplementary Figure S8. Venn diagram showing the overlap in terms of molecular formulae in the three different FT-MS datasets (A, red, Rio Negro and two tributaries, Simon et al., 2019; B, yellow, Rio Negro and lakes alongside the river, Gonsior et al., 2016; and C, blue, whitesand area dataset, this study).Panels show different sets of molecular formulae: a) whole set of molecular formulae; *asterisk: 70 CHONS formulae not included in panels c and d).b) Only formulae without Nitrogen or Sulfur atoms, c) Only formulae containing one or two N atoms, d) Only formulae containing a Sulfur atom.



Supplementary Figure S9. Result of the cluster analysis taking into account all formulae present in each measurement of the three datasets under study across the mass range m/z 180 – 800 (number of formulae = 7500). Ion abundance information was omitted to reduce instrument-specific effects (such as tuning, ionization, etc.). Clusters: a) Samples from this study; b) data from Gonsior et al. (2016); c) data from Simon et al. (2019). Clustering was conducted in R Studio by function vegdist (with Bray-Curtis dissimilarity) of vegan package in R Studio and hclust (with Ward linkage, “ward.d2”) of stats package. Groundwater/ headstream DOM from whitesand areas (a) is most dissimilar from river DOM (b, c). Spatial variability among sample sets is more pronounced in the groundwater dataset (a), probably due to lower heterogeneity in aquatic settings (mixing, etc.). However, even headwater streams differ strongly (RA, RC) from river samples (b, c), hence suggesting compositional changes during downstream transport.

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

Data from blackwater samples of the two datasets (Gonsior et al., 2016; Simon et al., 2019) were merged with data from both whitesand areas to assess the degree of overlap between datasets and environments. Initial formula numbers were 5119 (this study), 4958 (Gonsior et al., 2016) and 3561 (Simon et al., 2019). Scan ranges differed slightly (m/z 120 – 1000, 180 – 800, 150 – 800), same as the range of detected signals (m/z 120 – 801, 180 – 799, 154 – 661), sample flow rates (7 µl/min, 2, 2), accumulation/ inlet times (100 ms, 200-500, 200), scan number (300, 500, 500) and presumably C concentration during electrospray ionization (ESI) in negative mode (20 mg/L in this study and in Simon et al. 2019, but not clearly stated in Gonsior et al. 2016). Similar to the chosen ionization mode (ESI negative), resolution at m/z 400 was in the same order of magnitude (480k, 500k, 500k). Besides site and lab effects, accumulation time has to be regarded as the main factor of variation under these otherwise similar measurement conditions (Hawkes et al., 2016; Simon et al., 2018).

We used the whole lists of detected formulae across Rio Negro samples (formulae detected at least once) of both studies, yielding a total of 24 additional blackwater DOM measurements (Gonsior et al. (2016): 18 measurements of ten sampling stations; Simon et al. (2019): six measurements from six sampling stations including two Rio Negro tributaries). We used the data as downloaded. To bring datasets into comparable format, we removed 1) formulae detected below m/z 180 and above m/z 800 (to account for different scan ranges), 2) formulae containing P atoms or two S atoms or three to four N atoms (32 formulae with N3 excluded in Gonsior et al. 2016; 57 P, two S2 and twelve N3-4 formulae excluded in Simon et al. 2019). The remaining lists of each dataset were then merged by molecular formula and compared by Venn diagrams (overlap in terms of formula populations). Individual samples were compared by cluster analysis.

For the latter, the data were compiled to one crosstab and transformed to presence/ absence format. The clustering was achieved through combination of function vegdist of R package vegan (with Bray-Curtis dissimilarity) and hclust of R package stats (with Ward linkage for agglomeration, “ward.d2”).

The merged master list contained 7500 molecular formulae that represent an updated inventory of the Rio Negro watershed DOM spectrum (Data Set S2). A Venn analysis of the whole dataset revealed a common set of 2091 formulae (Figure S8) and major numbers of unique formulae in each dataset (common: ~28% of all formulae; specific to this study: 22%; spec. Gonsior et al. 2016, ~21%; spec. Simon et al. 2019, ~6%). The similarity in CHO formulae was a little higher compared to the total set of formulae (35%; 16%; 21%; 3%) whereas in terms of the CHNO formulae, similarity was lower than based on the total set (24%; 24%; 25%; 5%). Clear differences among sample sets became evident in case of CHOS formulae, with no single formula being part of all three sets and most sulfur formulae being found in dataset C (whitesand data set).

The comparison of datasets revealed that nitrogen- and especially sulfur containing formulae clearly differentiate the three datasets considered in this study. Although both types of formulae may be affected by anoxic conditions, and also by changes in the connectivity of riparian systems (Boye et al., 2017; Peyton Smith et al., 2017; Lynch et al., 2019), such differences may also be due to instrumental effects as heteroatom-containing formulae are harder to resolve and are often detected only at low ion abundance. Sulfur-containing formulae may also originate from contamination, by e.g. sulfonic acids.

The differentiation in terms of CHO, CHNO and especially CHOS formulae was also revealed by cluster analysis based on presence and absence of formulae in individual samples

(Figure S9). In general, the three datasets were clearly separated, and river samples were more similar to another than to any type of WSE-DOM. Although covering large spatial gradients, the both river datasets were strongly uniform in their molecular composition as compared to the soil water samples which showed stronger variation at a much narrower spatial scale. Samples from a suite of channels separated by river islands (cluster b; Gonsior et al. 2016), were astonishingly similar to each other besides the large spatial extent covered, as compared to Rio Negro and tributary samples (cluster c; Simon et al. 2019) and headwaters (cluster a, this study) that were more dissimilar. This likely reflects a more homogenized aquatic DOM pool as compared to soil environments (Kellerman et al., 2015; Lynch et al., 2019).

Supplementary Material References

Boye K., Noël V., Tfaily M. M., Bone S. E., Williams K. H., Bargar J. R. and Fendorf S. (2017) Thermodynamically controlled preservation of organic carbon in floodplains. *Nat. Geosci.* **10**, 415–419.

D’Andrilli J., Cooper W. T., Foreman C. M. and Marshall A. G. (2015) An ultrahigh-resolution mass spectrometry index to estimate natural organic matter lability. *Rapid Commun. Mass Spectrom.* **29**, 2385–2401.

Dansgaard W. (1964) Stable isotopes in precipitation. *Tellus* **16**, 436–468.

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

Hawkes J. A., Dittmar T., Patriarca C., Tranvik L. and Bergquist J. (2016) Evaluation of the Orbitrap Mass Spectrometer for the Molecular Fingerprinting Analysis of Natural Dissolved Organic Matter. *Anal. Chem.* **88**, 7698–7704.

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

Kellerman A. M., Dittmar T., Kothawala D. N. and Tranvik L. J. (2014) Chemodiversity of dissolved organic matter in lakes driven by climate and hydrology. *Nat. Commun.* **5**, 3804.

Kellerman A. M., Kothawala D. N., Dittmar T. and Tranvik L. J. (2015) Persistence of dissolved organic matter in lakes related to its molecular characteristics. *Nat. Geosci.* **8**.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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