Reviewer 1

Review of Biogeosciences Discuss. Doi: 10.5194/bg-2016-72

This manuscript describes an interesting study combining EEMS and FT-ICR-MS analysis of SPE-extracted DOM samples from a black-water, a white-water, and a clear-water river within the Amazon basin. As such it provides one of the most detailed compositional studies of the DOM within a water system to date and provides key information to add to the large body of work on bulk organic matter characteristics within the Amazon Basin. The topic and general scientific approach make the study a good one for publication in this journal. The problem with the manuscript right now is that it appears to have been written assuming that the reader will know all the details of the analytical techniques and multivariate analyses. Key information is either missing or presented but not explained.

General comments: In Section 2.2., the limits of detection, and a measure of precision and accuracy need to be given for each technique. In section 2.3, the authors need to tell the reader whether positive or negative ionization is used and what solvent the samples were brought up in and what the mobile phase was, if there wasn't direct injection of the sample. The details promised in the supplementary file are not present. In Figure 1 (representative FT-ICR-MS spectra) and in Figure S2, the reader needs to know which samples these were, mainstem or some other location. Perhaps the stations in Figure S1 could be given individual names or numbers and these could be used in the captions to Figure 1 and S1? Three panels in Figure 3 have –KMD/z* as the y axis but the reader is never told what this means (I assume a Kendrick mass defect factor) nor are these panels discussed anywhere in the text or supplemental section. Finally, the description of the heat maps needs to be revised for clarity. I tried very hard through multiple readings of the text (with the figures printed out for quick reference) to understand exactly what was being correlated in each figure but was unable to do so. Could there be some guidance along the x and y axes and more description in the text of the technique itself as well as the results?

Limit of detection and the standard deviations for each method was added in section 2.2

Additional information about the FT-ICR-MS technique was added: "*All SPE samples were analyzed using negative mode electrospray ionization and a Bruker Solarix 12 Tesla FT-ICR-MS located at the Helmholtz Zentrum Munich, Germany. Details about the FT-ICR-MS analyses used in this study have been described previously (25, 30). Briefly, methanolic samples were diluted 1:20 with methanol and then directly inject into the electrospray at a flow rate of 120 µL min^{-1}. Five hundred scans with a time domain of 4 megawords were averaged and the averaged spectra were post-calibrated using a list of known DOM internal calibrants."*

Additional information were added to Figure 1. Sample FT-ICR-MS spectra were from the mainstem of the rivers.

Additional information about the visualization tools are given in the method section: "*Van Krevelen diagrams (41) were used to visualize the elemental ratios of unambiguously assigned molecular formulas. Kendrick plots (42) are also useful to determine members of homologous series, but we used a modified Kendrick plot, where the Kendrick mass defect (KMD) is divided by another independent parameter z* (33) to describe homologous series and molecular formulas only spaced by CH2. This ratio of KMD divided by z* (KMD/z*) enabled the unambiguous determination of homologous series* *and an enhanced visualization (much better resolution between homologous series). Additional details about this approach were previously described (43).”*

Additional explanations are now added to the method section and Figure 5 and S6 were revised and the color legend was added, as well as what the horizontal and vertical lines are.

Specific comments:

1. The superscripts throughout this manuscript appear to have been lost in a formatting step.

Corrected

2. Page 2, line 12, “has” should be “have”

Corrected

3. P.3, line 11, “exists” should be “exist”

Corrected

4. P. 4, line 7, should read “Small (40-mL) aliquots”

Corrected

5. P. 4, line 18. Which convention should be used in this journal for five hundred thousand, the decimal or comma form?

Corrected to use comma as separator

6. P. 4, line 30. “One hundred microliters”?

Corrected

7. Figure 1 is mislabeled in its caption as Figure 2.

Corrected

8. P. 6, lines 3-6. This sentence is trying to say too many things and I cannot follow the points. Please reword. Perhaps start by pointing out the formulae for your hydrogen deficient vs more saturated compounds

The sentence was simplified and it read now as follow: “*The Rio Negro mass spectrum, in comparison to the other areas, clearly showed much higher intensities of hydrogen-deficient m/z ions in the low and high molecular weight ranges (Fig. 1).”*

9. P. 6, line 13, “where” should be “were”

Corrected

10. P. 7, line 3, what does “disk-shaped” mean? You cannot know higher order folding of the molecules from MS data, can you? Or are you referring to a disk-shaped distribution of formulae in the van Krevelen?

The word disk-shaped was deleted. What was meant here is that these aromatics are planar and hence may undergo charge transfer if they get in close proximity.

11. P. 7, line 29, “close” should be “closely”

Corrected

12. P. 7, line 31, “high” should be “highly”

Corrected

13. P. 8, line 10, should read “long-wavelength absorbance” \

Corrected

14. P. 8, lines 8-13. I am not sure I am following this reasoning. Please clarify.

This statement was simplified to: ” *…areas, but presumably non-fluorescent aliphatic CHOS and*

CHNO compounds were indicative for the Madeira sampling area and which were outside the analytical window of the EEM-PARAFAC approach (Fig. 4)."

15. P. 8, lines 14-16. Please reword this sentence for better clarity.

The sentence was changed to: *"Our results indicated that classified Amazon water systems (black, white and clear water) were associated with the presence or absence of many regionally unique compounds. It remains an open……"*

16. P. 8, line 27, add "of" before "which"

Corrected

17. Please add information on the color scale and axis or variable labels for the heat maps.

The figures containing heatmaps were updated to include color legend and more details about horizontal (m/z values ot Fmax values) and vertical (samples) lines.

Reviewer 2

Anonymous Referee 2 Comments for Biogeosciences Discussions manuscript "Chemodiversity of dissolved organic matter in the Amazon Basin" MS number: bg- 2016-72

Overall: The manuscript by Gonsior and coauthors compares the DOM quality using traditional optical methods (EEM, PARAFAC) with more specialized FT-ICR-MS and statistical methods in three different water types in the Amazon Basin. They present some interesting data and it is exciting that they saw definite relationships between the optical and molecular methods and the unique chemical signals between the three rivers. I think this paper would be improved by some rearrangement and additional content and clarification, especially regarding which samples are in the confluence of the Amazon River and how that relates to removal of DOM through adsorption and/or coagulation, currently the data presented does not clearly support DOM removal.

Authors Response: It was not possible during the sampling to directly compare Rio Negro waters with Solemoes waters at the confluence east f Manaus, but we added results from previous studies that strongly support a removal of DOM from the Rio Negro River. Our data supports these earlier findings and gives the first molecular fingerprint of this DOM component that is presumably adsorbed or coagulated. It was not the aim of this study to understand in detail the mechanism of this removal but we felt that a reasonable explanation would be coagulation with suspended solids transported by the Solemoes and Madeira Rivers.

Specific Comments: For the most part, the sampling locations are divided into Rio Negro, Rio Madeira and Rio Tapajos. However, there are a couple of other location that are also presented e.g., Rio Jamari (abstract and tables), Solimoes River (discussion); additional information needs to be presented to give context to these locations or include the data under the three main location. Are all of the samples from the Tapajos River in the confluence? It is not clear what stations are part of the confluence and which samples are just part of the river stem or flooded lakes. This needs to be clarified throughout the manuscript, but especially in the results and the discussion since a main point of your paper is the removal of DOM upon the mixing of rivers. In the current version of the manuscript the data is not presented to clearly demonstrate that DOM is being removed.

Authors Response: Additional information was added to the method section to explain in more detail where the samples were collected. The removal of Rio Negro DOM after mixing with the Solemoes has been demonstrated before and it is only one part of this study to show which DOM component is unique for the Rio Negro and hence is likely the one removed after mixing with high suspended sediment rivers. We tried to make this point clearer throughout the manuscript and please see responses to specific comments below.

Pg 1, Line 21: Are all the samples from the Tapajos in the confluence? reword if this is not the case.

No. Some samples were taking directly from the Tapajos and adjacent flooded lakes. (3). All other samples were collected in the confluence but most showed a much higher contribution from the Tapajos when compared to the mainstem of the Amazon. Hence, the area was named "Tapajos Area". More details were added in the method section to clarify this in more detail.

Pg 2, Line 19: Is it the removal of CDOM and suspended sediments that cause the waters to be clearer or the lack of CDOM and low suspended particle load due to clear waters draining weathered areas? Please clarify, add context, and/or estimates for each if available.

The lack of CDOM and suspended solids is the reason for the clear appearance of the Tapajos. We clarified this in the text.

Pg 2, Lines 29-34: How do clear waters fit into this? Add information if available or add a statement to the effect that no information is available for clear waters.

We added a reference to explain the Tapajos clear water system in more detail.

Pg 2, Lines 29-34: Tie this paragraph into how it would relate to your study; link to changes in chemical compounds

We tried to improve this paragrapha and added a sentence: …“These specific physico-chemical properties of these three main types of waters in the Amazon Basin are expected to exhibit distinctly different organic matter signatures.”

Pg 3, Line 16: How many surface water samples were collected in the main stem of the river? Include range as you do for the lake samples. Specify how many and which samples were collected in the confluence vs the main stem.

Additional information about sampling locations was added to the method section

Pg 3, Line 23: Add temperature and duration that filters were combusted.

Added combustion temperature to method section

Pg 4, Line 1: Samples were kept cool when? During extraction, until frozen? Pleaseclarify.

Samples were kept on ice during the sampling period and later frozen. Details were added to the method section

Pg 5, Line 1: What was the pH of the original river waters? If not pH 4-5, why was this pH chosen instead of a more neutral pH? Did you observe a pH dependence for your EEMs if samples were run at pH 4-5 vs pH 7-8?

The pH was consistent for all samples after dissolving the dried SPE-DOM in Milli-Q water. A narrow pH range of all samples is important to not introduce a pH bias as shown in our previous publication (Timko, S. A., M. Gonsior, and W. J. Cooper. 2015. Influence of pH on fluorescent dissolved organic matter photo-degradation. Water Research 85: 266-274.)

Pg 7, Lines 3-7: The Solimoes River has not been mentioned before and no data are presented that show a decrease in DOM from the Rio Negro; the values for the flooded lakes and main river stem are almost identical for DOC. If the Solimoes River is just used as a comparison from another study, data needs to be presented that shows DOM removal occurred in the Rio Negro (DOC values and/or FT-ICR-MS figures from above and below the confluence).

Additional information from previous studies was added to make this point clearer. The Solimoes was unfortunately not sampled during our sampling campaign, but previous data strongly supports the removal of DOC from the Rio Negro after mixing with the Solemoes.

Pg 7, Lines 9-19: This paragraph would benefit from some reorganization and clarification to allow this paragraph to flow better and increase understanding. Move the sentence “Very low abundance: : :” (lines 11-12) after the discussion about Rio Madeira and/or provide more context for the sentence, especially related to the boreal lakes and link to the next paragraph “The Tapajos areas contained: : :” (lines 17-19). Simplify the sentence “The unique and diverse: : :” (lines 13-15), are these the same or different unique compounds from the first sentence or the other unique compounds with unknown origin? Provide context as to how growing soy beans could cause unique compounds.

The sentence of the comparison with boreal lake DOM was deleted, because it distracted from the findings. The paragraph was revised and a sentence was added at the end of the paragraph: "*A potential source of sulfur are sulfonates which may origin from daily care products (e.g. surfactants), but also from wetting agents in fertilizers.***"**

Pg 7, Line 23: How does Fmax3 relate to the Rio Negro? Fmax3 looks like it would also be a dominant component in Rio Negro in the example EEM in Fig. S2 and you state that Fmax 3 and 4 typically were correlated.

You are absolutely right. Fmax3 behaves like Fmax4. This was corrected in the text.

Pg 7-8, Lines 32-1: Unclear how the coagulation of analogous molecular ions fit into this paragraph. Suggest removing sentence.

The sentence was removed

Pg 8, Line 13-14: This sentence is vague and needs clarification. I agree that light attenuation would be different for the various sampling locations due to the differencesin the water characteristics, but how is this related to the unique compounds.

The sentence was clarified to: "*Our results indicated that classified Amazon water systems (black, white and clear water) were associated with the presence or absence of many regionally unique compounds.*"

Pg 8, Lines 14-16: Additional discussion needs to be made to support this statement. No data is presented that suggests removal or transport of FDOM on particles or that you observed a loss of FDOM in the study area. If additional data are given to support the adsorption/coagulation, this statement could be expanded upon to tie into the bigger picture of biogeochemical cycling.

Previously published literature values were added and an additional figure (S5) was added to the supplementary material to emphasize the much higher long-wavelengths EEM-PARAFAC components in the Rio Negro samples when compared to all other samples. The text was expended accordingly.

Pg 8, Lines 17-19: Please present evidence that the differences in the DOC concentration are a result of adsorption or coagulation besides just a difference in FT-ICR-MS. These three rivers seem to have very different geology in their headwaters that could also be responsible for the differences in DOC concentrations. This is especially relevant to the Rio Tapajos waters that should have a very low particle load.

Additional convincing literature data were added to the manuscript.

Pg 8, Line 22: Could you provide some data to support the removal of HMW compounds from this river? Possibly a comparison of EEMs or FT-ICR-MS from upstream (lower particle load) and downstream (higher particle load)

The text was clarified and links to the figures that already showed this in the FT-ICR-MS data are given. An additional Figure S5 was added to show the much enhanced intensity of long-wavelength absorbance and fluorescence in the Rio Negro samples.

Pg 9, Line 22-25: Additional information or clarification is needed for this section as stated previously regarding removal. The plants in the region may not vary substantially throughout the region but there is likely additional factors for the differences in DOC concentrations than just adsorption to mineral particles or coagulation with metals, likely a result of the geology in the different regions. Also no data is shown to support DOM removal through mixing.

Additional explanations were added to the manuscript

Pg 16: Table 2: Rio Jamari is only mentioned in this table and the abstract. Provide some context for this location, is it another white water river like Rio Madeira? Or was this what the river was called after confluence with the Amazon River?

The Jamari is a small high sediment load tributary to the Madeira but it is also flooded by the Madeira during the high water season. Some more details are given in the method section.

Pg 16: Table 2: It is not clear which data set is after the confluence with the Amazon River. Please provide clarification as Rio Tapajos has the same labeling (flooded lakes and main stem river) as the other rivers. Pg 21, Fig. 5: To increase the understanding for readers unfamiliar with heat maps, include color bar legend and labels for the various axes.

The table was clarified

Fig. S1: Are these stations for the flooded lake sites only or all sampling locations? Add text to caption for clarity. Can you provide additional information and make the map marker a different color for stations that were within the confluence with the Amazon River.

More explanation was added to the method section to make this clearer

Technical Comments: Introduction:

Pg 2, Line 9-10: Switch “clear waters” and “white waters” since that is the order you talk about them.

Corrected

Pg 2, Line 9: I suggest rewording the beginning of this paragraph to “Amazon tributaries vary in their coloration and opacity due to their origin and reactivity and have traditionally been classified as “black waters”, “white waters” and “clear waters”. These three water types play a continuing role: : :”

Corrected

Pg 2, Line 11: Suggest using a different word than “processing” and simplifying “such as through”

Changed wording and replaced “processing” to “transformation”

Pg 2, Line 12: Change “has” to “have”

Corrected

Pg 2, Line 23: Reword. Possibly to “High CDOM in black waters and suspended sediment concentrations in white waters limit light: : :”

Corrected

Pg 3, Line 2-3: Reword. “It is unclear how previous work: : :or microbial processes, reflect the authentic: : :”

Corrected

Pg 3, Line 10: Change “By applying” to “Using”

Corrected

Pg 3, Line 10: Correct “overall”

Corrected

Pg 3, Line 11: Simplify sentence by removing “moreover”

Corrected

Pg 3, Line 13: Change “issues” to “relationships” Materials and Methods: Pg 3, Line 16: Remove “by”

Corrected

Pg 3, Line 17: Check number of lakes sampled, in Fig. S1, n=10,n=10, n=9 for Madeira, Negro, and Tapajos Rivers respectively. Should it be “9-10 lakes”? Or correct Figure S1.

Corrected to 9-10 lakes

Pg 3, Line 20: Move “either”; “were filtered and solid phase extracted either immediately after collection: : :or within three hours”

Corrected

Pg 3, Lines 25-26: Reword sentence to “Formic acid was used instead of HCl to prevent possible chloride ion adduct formation: : :.”

Corrected

Pg 3, Line 28: Reword line 28 to include “pH 2, formic acid” from line 30, since this is the first occurrence of “acidified Milli-Q water”; if defined in line 28, do not need to include in line 30.

Corrected

Pg 4, Line 4-5: Is this the range for all samples and location? Reword: “DOC concentrations for all samples ranged from 3-10 mg L-1, so only 1 L of sample water was filtered through the SPE cartridges to prevent overloading the 1 g cartridges.”

Corrected

Pg 4, Line 8-9: Move manufacture’s name: “: : :performed by an automated FOSS® colorimetric flow injection analysis (FIA) system, according to the quality control guidelines recommended by the manufacturer” Ultrahigh Resolution Mass Spectrometry

Corrected

Pg 4: I would recommend rearranging this paragraph so you introduce what you did before providing explanation.

The paragraph was changed as suggested.

Pg 4, Line 15-16: Did your instrument achieve a mass accuracy of less than 0.2 ppm? If so please clarify. State what your instrument did rather than saying what is typically achieved.

Yes it did. This was corrected.

Pg 4, Line 25-26: Possibly move this sentence “All SPE samples: : :” to the beginning of the paragraph so the reader knows what you did prior to additional explanation. Excitation Emission Matrix Fluorescence and Parallel Factor Analysis:

Moved this sentence to the beginning

Pg 4, Line 29-30: Rearrange sentence: “CDOM was recovered almost quantitatively (>90

Corrected

Pg 4, Line 30: Add volume to sentence “XX hundred uL: : :” or add a transition, “SPEDOM samples were prepared by drying 100 uL of methanolic SPE-DOM: : :”

Corrected

Pg 5, Line 12-13: This sentence could be simplified to “The maximum intensities of the components (Fmax) for each sample was exported and used for all subsequent statistical analyses.”

Corrected

Pg 5, Line 25: Which “two normalized data sets” are you referring to?

More details were added.

Results and Discussion:

Pg 6, Line 2-6: Break into two sentences, “: : :and Rio Madeira. Conversely, at high m/z, e.g. NM 601: : :”

Corrected

Pg 6, Line 10: Add “however” “However, the intensity-weighted: : :” Since the decreasing trend was not the same for elemental ratios as for the DBE.

The trend was the same, but literature citations may have been confused with data. This was corrected!

Pg 6, Line 11: Replace ‘that’ with ‘the’ or ‘the same’, “did not follow the same decreasing trend: : :”

Corrected

Pg 6, Line 14-16: Be more specific with how the different formulae were different in the various water types, e.g. instead of ‘noticeably lower’ maybe use ‘approximately 50

More specific details were added

Pg 6, Line 22: Clarify that these 6118 molecular formulae are for all the samples from all the waters combined and link to the top 6 panels of Fig. 2 “All Amazon DOM signatures”

This was clarified in the text.

Pg 6, Combine paragraph (lines 18-24) and next paragraph (25-27) or tie them together better.

Corrected

Pg 6, Line 25: Remove “only”

Corrected

Pg 6, Line 25-26: Tie this sentence to the lower 6 panels of Fig 2, possibly by using “ubiquitous” instead of “common”.

Corrected

Pg 7, Line 3: The phrase “disk-shaped” is odd, is there a better word choice.

This was deleted for clarity.

Pg 7, Line 19: Add ‘found’, “signatures found in all” Pg 7, Line 24: Provide reference for conjugated _-systems.

Corrected

Pg 7, Line 29: add “to” “remarkably close to the results”

Corrected

Pg 8, Line 7: Change ‘However’ to ‘For example’ or reword. This sentence seems like it is in support of the previous discussion that the optical properties could be related to specific molecular characteristics.

This sentence was changed to: “…waters from various Amazon Basin areas, but presumably non-fluorescent aliphatic CHOS and CHNO compounds were indicative for the Madeira sampling area which were outside the analytical window of the EEM-PARAFAC approach (Fig. 4).”

Pg 8, Lines 21-22: Simplify sentence: “The DOC data were in agreement with the FT-ICR-MS and EEM-PARAFAC results: : :”

Corrected

Pg 8-9, Line 32-2: Simplify and reword. Possibly “One unique CHO cluster consisted of hydrogen-deficient (low H/C), but highly oxygenated (high O/C) molecular ions and had a strong positive correlation with all Rio Negro samples (Fig. S6). Similarly, the Fmax components 3 and 4 were highly correlated to a cluster of high molecular weight molecular ions likely aromatic in origin (Fig. 5 and S6).” Tie this back to lignin.

Paragraph was restructured and tied to a possible humification: “*One unique CHO cluster consisted of hydrogen-deficient (low H/C), but highly oxygenated (high O/C) molecular ions and had a strong positive correlation with all Rio Negro samples (Fig. S7). Similarly, the Fmax components 3 and 4 were highly correlated to a cluster of high molecular weight molecular ions likely aromatic in origin (Fig. 5 and S7). This unique DOM signature showed much higher O/C ratios when compared to known lignins and tannic acids and may resemble the result of humification and an increase in non-lignin aromatic structures and higher carboxyl group content (59).”*

Pg 9, Lines 6-9: Simplify the sentence “Components Fmax 3 and 4: : :” (lines 6-9) to “However, components Fmax 3 and 4 did not show any correlation with molecular ions in the CHNO pool, nor in the CHOS pool, supporting the supposition that these PARAFAC components were only derived from CHO molecules.” May fit better immediately after the rest of the discussion about Fmax 3 and 4, end of line 2.

Moved and corrected

Pg 9, Line 11: Remove comma and ‘was’.

Corrected

Pg 9, Line 16: Reword: “Overall, Fmax 1, 2, and always correlated together as did Fmax 3 and 4.”

Corrected

Pg 9, Line 21: Change ‘appearance’ to ‘compounds’.

Corrected

Pg 9, Line 27: Reference Table 2 and highlight the similarities between the river and flooded lake.

Referenced

Pg 9, Line 30-33: Split into two sentences. Pg 9, Line 34: Be specific about which methods. “: : :without FT-ICR-MS and statistical analyses”

Corrected

Tables: Pg 16: Table 2: Give number of samples collected for each location.

Added

Figures: Pg 17: Should be Figure 1

Corrected

Supplemental Figures: Rearrange supplemental figures so that they are numbered as they are listed in the text.

Corrected

Fig. S1: Rearrange the station order since you always talk about them in the order Rio Negro, Madeira, and Tapajos.

Corrected

Fig. S1: Since this is a supplementary figure, possible to make maps larger?

Maps were enlarged

Fig. S6: To increase the understanding for readers unfamiliar with heat maps, include color bar legend and labels for the various axes.

Legend added

Chemodiversity of Dissolved Organic Matter in the Amazon Basin

Michael Gonsior[1], Juliana Valle[2], Philippe Schmitt-Kopplin[3,4], Norbert Hertkorn[4], David Bastviken[5], Jenna Luek[1], Mourad Harir[4], Wanderley Bastos[6], and Alex Enrich-Prast[2,5]

[1]Chesapeake Biological Laboratory, University of Maryland Center for Environmental Science, Solomons, MD20688, USA
[2]Departamento de Ecologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 21941-901, Brazil
[3]Analytical Food Chemistry, Technische Universität München, Freising-Weihenstephan, 85354, Germany
[4]Research Unit Analytical BioGeoChemistry, Helmholtz Zentrum Muenchen, Neuherberg, 85764, Germany
[5]Department of Thematic Studies - Environmental Change, Linköping University, Linköping, 581 83, Sweden
[6]Laboratory of Environmental Biogeochemistry, Universidade Federal do Rondônia, Rodovia, 76801-974, Brazil

Correspondence to: Michael Gonsior (gonsior@umces.edu)

Abstract. Regions in the Amazon Basin have been associated with specific biogeochemical processes, but a detailed chemical classification of the abundant and ubiquitous dissolved organic matter (DOM), beyond specific indicator compounds and bulk measurements, has not yet been established. We sampled water from different locations in the Negro, Madeira/Jamari and Tapajós River areas to characterize the molecular DOM composition and distribution. Ultrahigh resolution Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) combined with excitation emission matrix (EEM) fluorescence spectroscopy and Parallel Factor Analysis (PARAFAC) revealed a large proportion of ubiquitous DOM but also unique area-specific molecular signatures. Unique to the DOM of the Rio Negro area was the large abundance of high molecular weight, diverse hydrogen-deficient and highly oxidized molecular ions deviating from known lignin or tannin compositions, indicating substantial oxidative processing of these ultimately plant-~~derrived~~derived polyphenols indicative of these black waters. In contrast, unique signatures in the Madeira/Jamari area were defined by presumably labile sulfur and nitrogen-containing molecules in this white water river system. Waters from the Tapajós ~~confluence area~~mainstem did not show any substantial unique molecular signatures relative to those present in the Rio Madeira and Rio Negro, which implied a lower organic molecular complexity in this clear water tributary~~.~~, even after mixing with the mainstem of the Amazon River. Beside ubiquitous DOM at average H/C and O/C elemental ratios, a distinct and significant unique DOM pool prevailed in the black, white and clear water areas that were also highly correlated with EEM-PARAFAC components and define the frameworks for primary production and other aspects of aquatic life.

1 Introduction

With an average of about 200,000 m^3/sec and ~20% of the global freshwater discharge, the Amazon River transports larger amounts of riverine freshwater into the ocean than the next seven biggest river systems on Earth combined. DOM processed and transported within the Amazon Basin is therefore of global importance as 32-36 Tg of organic carbon (of which 60-70 % is dissolved) are exported seaward of Óbidos annually (1, 2). Repeated attempts have been made to describe the origin and fate of particulate and dissolved organic matter throughout the Amazon River and its tributaries using a variety of

Formatted: Superscript

methodologies. The sources of riverine OM have been identified through isotopic analysis of δ^{13}C, δ^{18}O, and δ^{15}N (3-8). Humic acids, fulvic acids, and lignin concentrations have also provided information on the OM sources (3, 4, 9) whereas labile DOM components such as saccharides and amino acids (10), as well as bacterial ~~-~~consumption (11) and biological and photo-degradation observations (12) have provided information on the OM reactivity and bioavailability. ^{13}C NMR and optical spectroscopy have been used to understand bulk characteristics and functional groups of Amazon organic matter (OM) (7, 13, 14) and OM age has also been analyzed using δ^{14}C measurements (15), but a detailed molecular understanding has not yet been established for large tributaries of the Amazon River Basin.

Amazon tributaries vary in their coloration and opacity ~~and~~ due to their origin and reactivity and have traditionally been classified as “black waters”, “white waters” and “clear waters” ~~and “white waters”~~ (16). These three water types ~~are a function of their origin and environmental reactivity and also~~ play a continuing role in the ~~processing~~transformation of OM, ~~such as through~~due to mediating light availability for aquatic life or photoreactivity. Black waters are influenced by chromophoric DOM (CDOM) and ~~has~~have low particulate mineral content (16, 17). It has been suggested that drainage areas of black water systems are characterized by moist, acidic, hydric soils that allow for leaching of terrestrially derived plant matter, like lignins, tannins and other plant materials that also contribute to the CDOM (17). White waters make up ~2/3 of the Amazon Basin (16, 18), exhibit low DOM levels and are less acidic (pH 6.6 ± 0.2) when compared to the black waters of the Rio Negro (pH 4.5 ± 0.9) (19). The “white” color reflects a high mineral particle load due to drainage and erosion of calcic sedimentary deposits (20) originating largely from the Andes mountains. Clear waters drain kaolinite clays (17) and contain high concentrations of iron and aluminum oxides~~,~~ that may adsorb humic acids (21). Because of removal of CDOM and suspended sediments~~,~~ prior to the water entering the mainstem of these ~~waters~~clear water Amazon tributaries, the CDOM and suspended particle levels are ~~not as~~typically very low in these systems. As a result, clear waters are less light ~~limited~~limiting and can support higher phytoplankton biomass, if the generally low nutrient levels are elevated (1, 16, 22).

High precipitation in combination with low infiltration of flood plain soils produce rapid overland and shallow subsurface flow responses to precipitation events across the Amazon Basin which carry particulate organic matter (POM) and DOM to the river from the surrounding river corridor (6, 23).

High CDOM in black waters and suspended sediment concentrations in ~~black and~~ white waters~~, respectively,~~ limit light and therefore the autochthonous production of organic matter (22); accordingly, allochthonous inputs dominate the organic matter pool (9, 10, 15). In clear waters, light is abundant (22) but nutrients are limited and as a result, OM is still expected to be influenced by allochthonous input. However, agriculture and urbanization along clear waters can supply additional nutrients and therefore increase autochthonous OM production with potential consequences for the DOM pool. These specific physico-chemical properties of these three main types of waters in the Amazon Basin are expected to exhibit distinctly different organic matter signatures.

High bacteria counts (~109 L^{-1}) have been observed in black and white waters with peaks during annual floods (11) when terrestrial-derived DOM enters the river (8). The biodegradable DOM appeared to be rapidly removed from these waters causing bacterial carbon limitation (11). Clear waters showed phosphorus limitation during high water periods, but also

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

carbon limitation due to low carbon quality (24). Younger, biologically available organic matter is preferentially respired upstream and the apparent δ^{14}C age of POM increases downstream (11, 15). Therefore, a certain proportion of the DOM likely reflects the recalcitrant behavior after microbial metabolism.

The unique DOM environments found within the Amazon Basin, which are ultimately the drivers of aquatic life, have yet to be resolved on a fine scale. It is unclear ~~the role that~~how previous work on bulk DOM, optical properties, a few specific target compounds, isotopes, or microbial processes, reflect ~~on~~ the authentic chemical diversity intrinsic to the complex OM found in black, white, and clear waters. Therefore, we investigated the chemodiversity of these three water types across the Amazon Basin by employing non-target FT-ICR-MS interfaced with soft ionization electrospray (ESI), which has enabled the characterization of thousands of individual molecular ions in complex DOM mixtures (~~24-32~~25-33). By combining this technique with advanced optical characterization, excitation emission matrix (EEM) fluorescence spectroscopy and Parallel Factor Analysis (PARAFAC), we assess similarities and differences in DOM composition between different waters of the Amazon Basin.

~~By applying~~Using these techniques, we attempted to answer a) ~~What~~what is the ~~overal~~overall chemodiversity of DOM in the Amazon Basin, and ~~moreover,~~ do distinct differences in DOM composition ~~exists~~exist between major Amazon tributaries and flooded area? b) How well do the simple optical properties represent the overall molecular composition of DOM as described by FT-ICR-MS in tropical ecosystems? Satisfactory comprehension of these ~~issues~~relationships would have large implications for the understanding of aquatic food webs and also for predicting further transport and processing of DOM in the Amazon.

2 Materials and Methods

Water samples from the mainstem of the Madeira~~,~~ (and its small tributary Rio Jamari), Negro and Tapajós River ~~surface water samples~~ were collected in ~~duplicate by~~duplicates using 1 L pre-combusted Pyrex glass bottles. The bottles were filled in the main stem of the river just below the surface. In addition, ~~8~~9-10 lakes, that were flooded by the individual rivers at the time of sampling, were sampled in the same manner near the cities of Santarém (confluence of Tapajós and Amazon River: Rio Tapajós Area), São Carlos (Rio Madeira Area) and Novo Airão (Rio Negro Area) (Fig. S1). The sampling sites in the Rio Tapajós Area included samples from flooded clear water lakes adjacent to the mainstem of the Rio Tapajós and flooded lakes that were located after the confluence of Rio Tapajós and the mainstem of the Amazon River. However, the sampling sites after the confluence were still dominated by Rio Tapajós water. The collected samples were ~~either~~ filtered and solid phase extracted either directly after collection (aboard river boats) or within three hours after collection, when smaller boats were used for sampling.

Formatted: Superscript

Formatted: Font: Not Bold

2.1 Solid-Phase Extraction

All 1 L water samples were filtered through pre-combusted (500 °C) Whatman GF/F glass fiber filters, acidified to pH 2 by using high purity formic acid (98%) and subsequently solid-phase extracted (SPE). The used SPE procedure has been previously described in detail (~~33~~34), but it was modified ~~by using~~and formic acid was used instead of ~~hydrochloric acid (~~HCl~~). The reason behind this choice of acid, was that the remaining chloride ions in HCl may cause~~ to prevent possible chloride ion adduct formation in the ionization source of the FT-ICR-MS (~~34~~35). Briefly, SPE-cartridges (Agilent, Bond Elut PPL, 1g resin), were activated with methanol (Chromasolv, Sigma Aldrich), rinsed with acidified ultrapure water (Milli-Q, pH 2, formic acid) and the acidified water samples were gravity-fed through the cartridges at a flow rate of ~10 mL~~/~~ min^{-1}. Subsequently, the cartridges were rinsed again with acidified Milli-Q water ~~(pH 2, formic acid)~~ to remove remaining sample solution, dried and eluted with 10 mL high purity methanol (Chromasolv, Sigma Aldrich) into pre-cleaned 40 mL amber glass vials. ~~Samples~~Methanolic samples were then kept ~~cool~~on ice during the 2 weeks sampling period and later frozen at -18 °C. Methanolic extracts are stable when kept at -18 °C for extended periods of time (~~35~~336). The adsorption efficiency of the used SPE method was in general 90-98 % of the chromophoric DOM (CDOM) (measured by EEM fluorescence) and 60-70 % of the dissolved organic carbon (DOC). All samples were run in duplicates. ~~The~~ DOC concentrations ~~between~~for all samples ~~varied~~ranged between 3-10 mg L-1 ~~and hence~~, so only 1 L of ~~each~~ sample water was ~~used to make sure that~~run through the ~~used~~SPE cartridge to prevent overloading the 1 g ~~SPE cartridges were not overloaded~~PPL resin.

2.2 Nutrients, chlorophyll, DOC, total dissolved nitrogen (TDN) and total dissolved phosphorus (TDP) analyses.

Small (40 mL) aliquots of each water sample were filter~~-~~-sterilized (0.2 µm, Whatman GD/X Cellulose acetate filters) and stored on ice prior to analysis. Total dissolved nitrogen (TDN) and phosphorus analyses were performed by an automated FOSS® colorimetric flow injection analysis (FIA) system, ~~manufactured by FOSS®,~~ according to the quality control guidelines recommended by ~~FOSS~~the manufacturer. The detection limit for TDN and total phosphorus was 5 µg L^{-1} and standard deviations ranged between 0.0012 and 0.0023 mg L^{-1} and between 0.0015 and 0.0021 mg L^{-1}, respectively. DOC was analyzed in triplicates by oxidation with sodium persulfate in a titanium oven under high temperature and high pressure by using an automatic carbon analyzer (InnovOx Sievers TOC Analyzer~~). Chlorophyll a was extracted and analyzed according to a previously published procedure (36) and analyzed using a Turner Designs fluorometer (Model Trilogy).~~) with a detection limit of 0.05 mg L^{-1} and standard deviation between 0.01 and 0.02 mg L^{-1}. Chlorophyll a was extracted and analyzed according to a previously published procedure (37) and analyzed using a Turner Designs fluorometer (Model Trilogy) with a detection limit of 0.01 mg L^{-1}.

Formatted: Font: Not Bold, Font color: Black, Border: : (No border)

2.3 Ultrahigh Resolution Mass Spectrometry

All SPE samples were analyzed using negative mode electrospray ionization and a Bruker Solarix 12 Tesla FT-ICR-MS located at the Helmholtz Zentrum Munich, Germany. Details about the FT-ICR-MS analyses used in this study have been described previously (25, 30). Briefly, methanolic samples were diluted 1:20 with methanol and then directly inject into the electrospray at a flow rate of 120 µL min^{-1}. Five hundred scans with a time domain of 4 megawords were averaged and the averaged spectra were post-calibrated using a list of known DOM internal calibrants. A mass accuracy of less than 0.2 ppm deviation from the actual mass ~~is often~~was achieved (less than the mass of an electron~~) by this ultrahigh resolution technique.~~). Multiple charged ions in flow injection FT-ICR-MS analysis may occur dependent on the solution concentration and electrospray condition parameters (~~37~~38): however, in our analysis we only encountered singly charged ions as described previously (~~38~~39). The achieved mass resolution was in routine full scan 500~~.~~,000 at *m/z* 400 and exact unambiguous molecular formulae were assigned to the observed molecular ions up to a *m/z* of 800 Da (~~39~~40). Negative mode electrospray ionization (ESI) typically generates several thousands of different *m/z* ions, but this ionization technique is largely biased toward organic acids, because of their high ionization efficiencies. For example, alcohols and saccharides do not ionize efficiently in electrospray and hence signals of these compound classes are largely lost in complex DOM mixtures analyzed by FT-ICR-MS. On the other hand, DOM contains a large diversity of highly polar and polyfunctional easily ionizable compounds (e.g. organic acids), operationally also classified as fulvic and humic acids.

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

~~All SPE samples were analyzed using a Bruker Solarix 12 Tesla FT-ICR-MS located at the Helmholtz Zentrum Munich, Germany. Details about the FT-ICR-MS analyses used in this study have been described previously (24, 29) and additional details including explanations of visualization tools are given in the supplementary material.~~

Van Krevelen diagrams (41) were used to visualize the elemental ratios of unambiguously assigned molecular formulas. Kendrick plots (42) are also useful to determine members of homologous series, but we used a modified Kendrick plot, where the Kendrick mass defect (KMD) is divided by another independent parameter z* (33) to describe homologous series and molecular formulas only spaced by CH2. This ratio of KMD divided by z* (KMD/z*) enabled the unambiguous determination of homologous series and an enhanced visualization (much better resolution between homologous series). Additional details about this approach were previously described (43).

2.3 Excitation Emission Matrix Fluorescence and Parallel Factor Analysis

~~The used SPE method~~ CDOM was recovered ~~CDOM~~ almost quantitatively (>90~~%)~~ %) using the described SPE method and ~~in order to directly compare~~enabled a direct comparison of FT-ICR-MS results and optical properties of Amazon DOM~~, we also used~~. SPE-DOM samples were prepared for ~~the determination of~~ optical ~~properties. Hundred~~analyses using 100 µL of the methanolic ~~SPE-DOM sample~~extract that was completely dried under pure nitrogen; re-dissolved in 10 mL Milli-Q water and further diluted with Milli-Q water (1:20). The pH of the solutions of all re-dissolved samples ranged between 4-5 pH units. Optical properties are highly dependent on pH and hence a narrow pH range was beneficial to accurately compare samples. Excitation emission matrix (EEM) fluorescence measurements of dried and re-dissolved SPE-DOM were measured using a temperature-controlled Jobin Yvon Aqualog fluorescence spectrometer. The emission was recorded over the range from 200-600 nm for excitation wavelengths ranging from 240-600 nm at 3 nm intervals. Fluorescent intensities were Raleigh scattering corrected, inner filter corrected by using the absorbance data and normalized to a STARNA quinine sulfate fluorescent standard of 1 ppm concentration. The statistical Parallel Factor Analysis (PARAFAC) of fluorescence data (~~40~~44) was applied in this study to the EEM data set by using drEEM, which was developed in Matlab and utilizes the N-way toolbox (~~41~~45). Several models were tested and a 5 component model was developed on the normalized data (Fig. ~~S3~~S2). This model showed the best results in terms of separation between components, residuals and core consistency. Normalization was reversed prior to split-half validation of 6 subsets of the 5 component model (~~41~~45). The 5 component model was split-half validated and explained 99.86 % of the variability and the results of the split-half validation are given in supplementary information (Fig. ~~S4~~S3). The ~~Matlab procedure drEEM allowed exporting the data of the~~ maximum intensities of the components (Fmax) ~~to be able to also take into account fluorescence intensities of individual samples. Fmax scores were~~for each sample was exported and used for all subsequent statistical analyses.

2.4 Multivariate Statistical Analysis~~.~~

Data mining and the application of multivariate statistics is increasingly important to be able to analyze very complex data sets. Examples of such multivariate approaches are hierarchical cluster analysis (HCA) and principle component analysis (PCA) that have been recently applied to ultrahigh resolution mass spectrometry (~~42-44~~46-48) and EEM-PARAFAC (~~45~~49). In this study, PCA and HCA were applied to mass spectrometry-based data sets of all collected spectra and their exact mass lists and intensities. The duplicate samples were first averaged and resulted in up to 16,000 variables (*m/z* ions). A data matrix was compiled of all averaged samples and mass lists, where *m/z* ions were matched in a narrow 0.2 ppm error window. This matrix was then normalized by subtracting the average value from each data point and dividing it by the standard deviation (~~46~~50). The resulting FT-ICR-MS mass list data sheet was then converted into a resemblance matrix by using Spearman Rank correlations and used to create hierarchical clusters. The same data set was also used in the PCA analysis of FT-ICR-MS data. ~~-~~A similar approach was undertaken for the EEM-PARAFAC data set.

In an additional analysis, the two normalized data sets (FT-ICR-MS and EEM-PARAFAC) were combined to be able to determine hierarchical clusters on the variables (*m/z* ions and their intensities and EEM-PARAFAC Fmax values~~).~~) and to create a microarray. This approach is in analogy to genetic data, where the expression levels of large numbers of genes can be simultaneously visualized. In our case, molecular formulas and their associated *m/z* ion intensities would be in analogy to specific DNA sequences and concentrations. This approach enabled the generation of heat maps (software: TM4-Multi Experiment Viewer) and hierarchical clusters of all molecular formula assigned to *m/z* ions and also which *m/z* ions or molecular formula co-varied with EEM-PARAFAC Fmax values. This approach successfully depicted hierarchical clusters of specific classes of ~~m/z ions~~molecular formulas with distinct and confined chemodiversity as expressed in van Krevelen

Formatted: Font: Bold

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

diagrams within the CHO, CHNO and CHOS pools (Fig. ~~S6), and revealed~~S7). Hence, subsets of ~~m/z ions~~molecular formulas that correlated well with EEM-PARAFAC components~~.~~ were depicted.

3. Results and Discussion

Representative FT-ICR-MS spectra from each area are provided in Figure 1. ~~At low m/z values, the intensities of m/z ions at e.g. nominal mass (NM) 265 suggested a shift from more hydrogen-deficient to more saturated compounds in the order Rio Negro, Rio Tapajós and Rio Madeira, whereas at high m/z, e.g. NM 601, the intensities of m/z ions were highest in the Rio Negro sample and decreased in the same order (Fig. 1) suggesting a much larger contribution of high molecular weight compounds in the Rio Negro samples.~~The Rio Negro mass spectrum, in comparison to the other areas, clearly showed much higher intensities of hydrogen-deficient *m/z* ions in the low and high molecular weight ranges (Fig. 1). Further evaluation of these initial results by intensity-weighted parameters confirmed the trend from higher mass to lower mass DOM molecules in the order Rio Negro, Rio Tapajós and Rio Madeira (Tab. 1). For example, the intensity-weighted center of mass with assigned CHO formulae was 453 Da in Rio Negro, 428 Da in Rio Tapajós and 419 Da in Rio Madeira. Counts of double bond equivalents (DBE) and DBE-oxygen (DBE-O) ~~(24, 32, 47)~~ also followed the same decreasing pattern from Rio Negro, Rio Tapajós and Rio Madeira (Tab. 1). The intensity-weighted averaged oxygen to carbon (O/C) and hydrogen to carbon (H/C) elemental ratios did not follow ~~that~~the same decreasing trend and indicated that the displacement of higher molecular weight DOM in the Rio Negro towards lower molecular weight in the Madeira or Tapajós Rivers was not associated with a change in overall elemental ratios. Somewhat similar trends where observed for the molecular ions that represented nitrogen-containing compounds, but the numbers of assigned CHNO formulae were ~~noticeably~~approximately 30% lower in the Rio Negro samples. Similarly, ions with sulfur-containing formulae were of very low abundance in the Rio Negro, when compared to higher occurrence in the Rio Tapajós and even higher proportions in the Rio Madeira (Tab. 1).

Detailed molecular formula assignments of duplicate SPE-DOM samples from the Rio Negro area (black water), Rio Tapajós area (clear water) and Rio Madeira area (white water) in both the main stem of the river as well as flooded adjacent lakes revealed an overall molecular composition of Amazon DOM that was extremely diverse (Fig. 2). CHO formulae covered almost the entire area of chemically reasonable O/C and H/C ratios, but also CHNO and CHOS formulae showed extensive compositional variance. The chemodiversity of ~~DOM in the~~all Amazon ~~Basin~~DOM signatures was characterized by 6,118 assigned molecular formulae occurring in duplicate samples of which 43 % were of hetero-atomic nature, i.e. CHNO, CHOS and CHNOS formulae~~.~~

(Fig. 2, top 6 panels). Common molecular formulae across all samples referred to 47 % of CHO, 31 % CHNO, but~~ only~~ less than 2 % of CHOS molecular formulas (Fig. 2). However, distinct area-specific differences in relative abundances applied to many of the ubiquitous DOM *m/z* ions~~.~~ (Fig.2, lower 6 panels).

A simple computation of unique signatures (*m/z* ions) from each sampling area (Fig. 3) was combined with molecular formulae assignments and revealed that the Rio Negro area was characterized by unique, abundant and numerous high

Formatted: Font: Italic

Formatted: Font: Italic

molecular weight compounds in a rather confined area within the van Krevelen diagram, indicative of a large diversity of polyphenolic-type compounds with high O/C (0.5-0.8) and low H/C (0.4-0.7) ratios, presumably highly degraded tannins or lignins. A large complement of the assigned formulae exceeded the common O/C ratios of known lignin or tannic acid sub-units and it appeared that these specific Rio Negro polyphenols were highly enriched in oxygen, possibly resulting from microbial side chain oxidation or even aromatic ring opening, which would lead to the incorporation of additional carboxylic acids (~~48~~52).

A removal of these unique ~~and likely disk-shaped~~ high molecular weight ~~polyaromatic~~poly-aromatic DOM molecules by means of mineral adsorption or flocculation with iron (21) or aluminum oxides (~~49~~53) after mixing with the high sediment-load Solimões River is conceivable and corresponded to previous reports that between 4% (~~50~~54) and 40% (2) of the DOC was removed at the confluence of the Rio Negro and Solimões. ~~Investigations~~We were not able to sample directly the Solimões River and hence a direct comparison between FT-ICR-MS results from the Rio Negro and Solimões were not achieved. However, investigations of Amazon wetland hydrogeochemistry also found analogous preferential sorption of higher molecular weight DOM to sediments, resulting in an enriched low molecular weight (LMW) aliphatic DOM pool under high suspended solids conditions (7).

In contrast, the distinct signatures indicative of the Rio Madeira area were comprised of diverse CHNO and CHNOS compounds that may define the readily bioavailable and labile DOM pool, because of in part its low molecular weight, in particular for the sulfur bearing *m/z* ions. ~~Very low relative abundance~~ A specific CHOS pool of ~~DOS and DON molecular ions were commonly found in other Amazon areas (Fig. 3) and also in boreal lakes (25). A rather unique~~ very low H/C ~~ratio CHOS pool~~ ratios was also apparent in the Madeira area samples~~, but its origin remains unclear.~~. The unique and diverse CHNO and CHOS compounds observed in the Madeira area samples ~~may also indicate~~might have indicated anthropogenic influence as suggested in a recent study (~~51~~54) and this suggestion was supported by the relatively high population ~~close~~closely to the Madeira sampling area. The city Porto Velho is about 30 miles upstream of the sampling locations and the whole area is intensively used for agriculture and to grow mainly soy beans. A potential source of sulfur are sulfonates which may origin from daily care products (e.g. surfactants), but also from wetting agents in fertilizers.

Formatted: Font: Italic

The Tapajós area showed indicative aliphatic CHO signatures (high H/C ratios) in the van Krevelen diagram and some diverse nitrogen-containing molecular formulae, but this region was in general characterized by ubiquitous molecular signatures found in all investigated areas.

EEM spectra of SPE-DOM also showed distinct differences between each area (Fig. ~~S2~~S4), and a ~~high intensity~~highly intense long~~–~~-wavelength fluorescent peak in Rio Negro water samples was apparent~~.~~ (Fig. S4). A five PARAFAC components (Fmax1-5) model (Fig. ~~S3~~S2) was most adequate to explain the differences in the fluorescence data set and also captured the indicative fluorescence signal of the Rio Negro (~~Fmax4~~Fmax3 and 4) at high emission wavelengths~~.~~ (Fig. S5). These high emission wavelengths can either be explained by large complex conjugated π-systems (56) that would support the unique aromatic high molecular weight DOM characteristic of Rio Negro waters or charge transfer processes (~~52~~57), which also would require rather complex molecular assemblies.

Additional multivariate analysis, such as HCA and PCA, of the FT-ICR-MS and EEM-PARAFAC data of all samples also produced distinct separation between sampling areas (Fig. 4). It was gratifying that the EEM-PARAFAC PCA and HCA clusters matched remarkably close to the results of the FT-ICR-MS data. These statistical results suggested that many of the molecular changes associated with each sampling area might be associated with certain changes in the fluorescent DOM (FDOM). This at least should apply for the Rio Negro area, because FT-ICR-MS-derived molecular signatures showed high hydrogen deficient molecules indicative of aromatic structures. ~~In a previous study, analogous molecular ions found in a comparable region in the van Krevelen diagram were removed through coagulation in a water treatment plant in Sweden that also removed the majority of CDOM (53).~~ However, correlations between optical properties (EEMs) and FT-ICR-MS data are not a certain proof of a causal relationship.

Traditionally, the classification of Amazon Rivers was based on appearance and color of the water, and largely defined by CDOM (e.g. black water - Rio Negro versus clear water - Rio Tapajós) and its sediment load (e.g. white water - Rio Madeira, Solimões). Hence, optical properties were important indicators at least for black and clear water systems. Our results suggested that this classification based on optical properties might expand to include specific molecular characteristics of waters from various Amazon Basin areas~~. However,~~, but presumably non-~~light absorbing~~fluorescent aliphatic CHOS and CHNO compounds were indicative for the Madeira sampling area and ~~might have been responsible for some~~which were outside the analytical window of the ~~smaller differences seen in the HCA between the FT-ICR-MS and~~ EEM-PARAFAC ~~data~~approach (Fig. 4). Aromatic CHNO compounds, which were present in the Madeira area samples, ~~might~~presumably carry the ability to show long~~ -~~-wavelengths absorbance and fluorescence, even at relatively low molecular weight. At present, it remains unknown whether aromatic nitrogen heterocyclic compounds play an important role in FDOM. Our results indicated that ~~light attenuation was~~classified Amazon water systems (black, white and clear water) were associated with the presence or absence of many regionally unique compounds. It remains an open question what happens with the FDOM that is presumably adsorbed to the mineral phase or coagulated and become part of the particulate fraction and whether or not it is respired or transported to the Atlantic Ocean and eventually desorbed, or if it is added to the downstream sediments or to the seasonally flooded forest floor.

DOC concentrations in the Rio Negro (10.8 mg/L) and surrounding lakes (10.1 mg/L) were twice or more of that of Madeira and Tapajós area waters (1.9-5.8 mg/L) (Table 2), consistent with earlier observations ~~(1, 2, 9),~~with mean DOC concentrations in the Rio Negro of 12.7 mg/L, Madeira of 5.8 mg/L, Solimõesof 5.8 mg/L (not measured in this study) and the Tapajós of 4.5 mg/L (1, 2, 9). It was also previously stated that the Rio Negro does not produce a simple dilution effect at the confluence with the Solimões (2). The Solimões and in particular the Madeira River transport by far the highest load of suspended solids of all tributaries in the Amazon basin (2) and the suggested adsorption or more likely coagulation of DOC when mixed with high suspended sediment rivers~~.~~ is conceivable. However, a direct comparison between the Rio Negro and the Solimões at the confluence downstream of Manaus was not undertaken in this study.

High Rio Negro DOC concentrations may be indicative of watershed characteristics (17) or a lower mineral content (2, 7, ~~50~~54) compared to the other tributaries. The DOC data were in agreement with the observations resulting from FT-ICR-MS

(Fig.1 and 3) and EEM-PARAFAC analyses (Fig. S5) that indicated a removal of high molecular weight polyphenolic-like compounds from the Rio Negro waters, when compared to samples collected downstream or in other catchments. Total dissolved nitrogen (TDN) and total dissolved phosphorous (TDP) were low in all waters (Tab. 2), contributing to the observed low primary productivity despite the differences in light availability between the systems. A high DOM C:N ratio was observed in all waters consistent with previous findings (9, ~~54~~58).

To address the question of which *m/z* ions were correlated with each other and also with what specific EEM-PARAFAC Fmax components, both data sets were combined, normalized and correlated using Spearman Rank correlations (Fig. 5). All variables were statistically compared within the dissolved organic carbon (CHO), dissolved organic nitrogen (CHNO) and the dissolved organic sulfur (CHOS) pools (Fig. 5 and S6). Heat maps revealed clearly distinct hierarchical clusters in each of the CHO, CHNO and CHOS pools, indicative of a very similar behavior of variables within individual clusters. Remarkably, large clusters in the CHO pool were found which represented rather confined molecular ions with specific O/C and H/C ratios (Fig. S6). ~~For example, in analogy to the~~One unique ~~molecular ions found in the Rio Negro area (Fig. 3),~~ CHO cluster consisted of hydrogen-deficient (low H/C~~ ratios~~), but highly oxygenated (high O/C~~ ratio~~) molecular ions ~~were clustered together including~~and had a strong positive correlation with all Rio Negro samples (Fig. ~~S6~~S7). Similarly, the Fmax components 3 and 4 were highly correlated ~~with an analogously aromatic~~to a cluster of~~ molecular ions with~~ high molecular weight ~~(Fig.~~ molecular ions likely aromatic in origin (Fig. 5 and ~~S6).~~S7).. This unique DOM signature showed much higher O/C ratios when compared to known lignins and tannic acids and may resemble the result of humification and an increase in non-lignin aromatic structures and higher carboxyl group content (59).

Formatted: Font: Italic

In contrast, Fmax1,2 and 5 only correlated with a very few aliphatic molecular ions (Fig. 5 and S6), which were improbable fluorescent compounds itself, indicating an indirect correlation. Accordingly, correlations of Fmax 1, 2 and 5 were found within a distinct CHNO cluster, indicating that the fluorophores responsible for these PARAFAC components might have derived from heterocyclic or other aromatic nitrogen-containing molecules. ~~Components~~ However, components Fmax 3 and 4 did not show any ~~correlations~~correlation with molecular ions in the CHNO pool, nor in the CHOS pool, supporting the ~~suggestion~~supposition that these PARAFAC components were only derived from CHO molecules~~ and also no correlations between m/z ions and Fmax 3 and 4 were found in the CHOS pool~~.

The heat map and correlations of CHOS molecular ions (Fig. S6) reflected the already mentioned enrichment of CHOS compounds in the Madeira and Tapajós area samples and the absence of these signatures in the Rio Negro area. The unique highly hydrogen-deficient CHOS pool only found in the Madeira sampling area~~, was~~ also manifested in a distinct cluster in this analysis. The CHOS-based cluster that co-varied with PARAFAC components Fmax1, 2 and 5 showed mostly molecular ions with high O/C ratios, and a high probability to be aromatic indicated by a subset of formulae that showed H/C ratios below 1. In contrast to the CHO pool that correlated with Fmax 3 and 4, this specific CHOS pool was confined to the low molecular weight range. At this point, it appears that CHOS molecules also contributed to the fluorescence signals manifested in Fmax 1,2 and 5. ~~Thus far, it was shown that~~Overall, Fmax 1, 2 and 5 always correlated ~~always~~ together ~~and~~

~~so~~as did Fmax 3 and 4. A simple Spearman Rank correlation between the PARAFAC components confirmed the very strong correlations between these two clusters (Fig. ~~S5~~S7).

Here, we present direct evidence that Rio Negro waters contained a unique, high molecular weight and highly fluorescent DOM component that is neither present in the Rio Madeira nor in the Rio Tapajós/Amazon confluence. This Rio Negro FDOM was confirmed by FT-ICR-MS to be indicative of high molecular weight polyphenolic-like ~~appearance~~compounds. DOC concentrations were also the highest in the Rio Negro. There is no reason to suspect widely different DOC produced by plants among the regions, but differences in geology is expected to influence the DOC pool, which is in part reflected in the large differences in pH and ~~therefore~~ it is ~~more~~ likely that this specific Rio Negro DOM fraction is sensitive to adsorption to mineral particles or coagulation with metals and that this specific DOM can be rapidly removed from the water during mixing of major Amazon tributaries~~.~~ as suggested in previous studies (2, 54). Such a removal would also substantially reduce the molecular complexity of Amazon DOM as observed in this study.

The similarities in DOM quality within regions between lakes and rivers ~~presumably having different residence times,~~(Tab. 2) indicated that the DOM composition is relatively stable under certain sets of environmental conditions. On the other hand, as discussed above, changing conditions may rapidly change DOM composition, in turn affecting conditions for aquatic life. While the commonly measured spectral properties of the DOM indicated these overall patterns, some important details were missed by solely measuring optical properties~~ and distinction~~. Distinction between unique types of DOM with different origin, functions, and turnover times, such as the previously discussed CHNO and CHOS compounds, would not have been possible without ~~other methods.~~FT-ICR-MS and multivariate statistical analyses. Thus, for improved capacity to understand and predict the fates and functions of DOM, new methods, such as ultrahigh resolution FT-ICR-MS, allowed detailed DOM characterization and emphasized the importance of non-target high resolution techniques.

4. Acknowledgements

This research was supported by the Brazilian research agencies FAPERJ, CAPES and CNPq and by the German research agencies Alexander von Humboldt and DAAD and the Swedish Research agencies STINT and VR. Enrich-Prast, A. has additional support as a research fellow from CNPq and a Cientista do Nosso Estado from FAPERJ. This is contribution XXXX (to be filled in after acceptance of manuscript) of the University of Maryland Center for Environmental Science, Chesapeake Biological Laboratory. All data used in this publication will be provided upon request by the corresponding author.

References

1. Richey JE, et al. (1990) Biogeochemistry of Carbon in the Amazon River. Limnology and Oceanography 35(2):352-371.

2. Moreira-Turcq P, Seyler P, Guyot JL, & Etcheber H (2003) Exportation of organic carbon from the Amazon River and its main tributaries. Hydrological Processes 17(7):1329-1344.

3. Hedges JI, Hatcher PG, Ertel JR, & Meyers-Schulte KJ (1992) A comparison of dissolved humic substances from seawater with Amazon River counterparts by carbon-13 NMR spectrometry. Geochimica et Cosmochimica Acta 56(4):1753-1757.

4. Hedges JI, et al. (2000) Organic matter in Bolivian tributaries of the Amazon River: A comparison to the lower mainstream. Limnol. Oceanogr. 45:1449-1466.

5. Quay PD, et al. (1992) Carbon cycling in the Amazon River: Implications from the 13C compositions of particles and solutes. Limnology and Oceanography 37(4):857-871.

6. McClain ME, Richey JE, Brandes JA, & Pimentel TP (1997) Dissolved organic matter and terrestrial-lotic linkages in the central Amazon basin of Brazil. Global Biogeochemical Cycles 11(3):295-311.

7. Maurice PA, Cabaniss SE, Drummond J, & Ito E (2002) Hydrogeochemical controls on the variations in chemical characteristics of natural organic matter at a small freshwater wetland. Chem Geol 187(1–2):59-77.

8. Moreira-Turcq P, et al. (2013) Seasonal variability in concentration, composition, age, and fluxes of particulate organic carbon exchanged between the floodplain and Amazon River. Global Biogeochemical Cycles 27(1):119-130.

9. Ertel JR, Hedge JI, Devol AH, & Richey JE (1986) Dissolved Humic Substances of the Amazon River System. Limnology and Oceanography 31(4):739-754.

10. Hedges JI, et al. (1994) Origins and processing of organic matter in the Amazon River as indicated by carbohydrates and amino acids. Limnol. Oceanogr. 39(4):743-761.

11. Benner R, Opsahl S, ChinLeo G, Richey JE, & Forsberg BR (1995) Bacterial carbon metabolism in the Amazon River system. Limnology and Oceanography 40(7):1262-1270.

12. Amado AM, et al. (2006) Complementary pathways of dissolved organic carbon removal pathways in clear-water Amazonian ecosystems: photochemical degradation and bacterial uptake. Fems Microbiology Ecology 56(1):8-17.

13. Luciani X, et al. (2008) Tracing of dissolved organic matter from the SEPETIBA Bay (Brazil) by PARAFAC analysis of total luminescence matrices. Marine Environmental Research 65(2):148-157.

14. Patel-Sorrentino N, Mounier S, & Benaim JY (2002) Excitation–emission fluorescence matrix to study pH influence on organic matter fluorescence in the Amazon basin rivers. Water Research 36(10):2571-2581.

15. Townsend-Small A, Noguera JL, McClain ME, & Brandes JA (2007) Radiocarbon and stable isotope geochemistry of organic matter in the Amazon headwaters, Peruvian Andes. Global Biogeochemical Cycles 21(2).

16. Sioli H (1950) Das Wasser im Amazonasgebiet. Forschungen und Fortschritte 26(21/22):274-280.

17. Leenheer JA (1980) Origin and Nature of Humic Substances in the Waters of the Amazon River Basin. Acta Amazonica 10(3):513-526.

18. Junk WJ (1984) Ecology of the várzea, floodplain of Amazonian whitewater rivers. The Amazon, Monographiae Biologicae, ed Sioli H (Springer Netherlands), Vol 56, pp 215-243.

19. Duncan W & Fernandes M (2010) Physicochemical characterization of the white , black , and clearwater rivers of the Amazon Basin and its implications on the distribution of freshwater stingrays (Chondrichthyes , Potamotrygonidae). Pan-American Journal of Aquatic Sciences 5(3):454-464.

20. Horbe AMC, Queiroz MMD, Moura CAV, & Toro MAG (2013) Chemistry of waters of the middle and lower Madeira River and its main tributaries- Amazonas - Brazil. Acta Amazonica 43(4):489-504.

21. Philippe A & Schaumann GE (2014) Interactions of Dissolved Organic Matter with Natural and Engineered Inorganic Colloids: A Review. Environ. Sci. Technol. 48(16):8946-8962.

22. Costa MPF, Novo E, & Telmer KH (2013) Spatial and temporal variability of light attenuation in large rivers of the Amazon. Hydrobiologia 702(1):171-190.

23. Johnson MS, et al. (2006) Organic carbon fluxes within and streamwater exports from headwater catchments in the southern Amazon. Hydrological Processes 20(12):2599-2614.

~~24~~24. Farjalla, V. F., F. A. Esteves, R. L. Bozelli, and F. Roland. Nutrient limitation of bacterial production in clear water Amazonian ecosystems. Hydrobiologia 489: 197-205.

25. Gonsior M, et al. (2011) Characterization of dissolved organic matter across the Subtropical Convergence off the South Island, New Zealand. Marine Chemistry 123(1-4):99-110.

~~25~~26. Gonsior M, Schmitt-Kopplin P, & Bastviken D (2013) Depth-dependent molecular composition and photo-reactivity of dissolved organic matter in a boreal lake under winter and summer conditions. Biogeosciences 10(11):6945-6956.

~~26~~27. D'Andrilli J, et al. (2010) Comprehensive characterization of marine dissolved organic matter by Fourier transform ion cyclotron resonance mass spectrometry with electrospray and atmospheric pressure photoionization. Rapid Commun Mass Spectrom 24(Copyright (C) 2011 U.S. National Library of Medicine.):643-650.

~~27~~28. Tremblay LB, Dittmar T, Marshall AG, Cooper WJ, & Cooper WT (2007) Molecular characterization of dissolved organic matter in a North Brazilian mangrove porewater and mangrove-fringed estuaries by ultrahigh resolution Fourier Transform-Ion Cyclotron Resonance mass spectrometry and excitation/emission spectroscopy. Mar. Chem. 105(1-2):15-29.

~~28~~29. Ward ND, et al. (2013) Degradation of terrestrially derived macromolecules in the Amazon River. Nature Geosci 6(7):530-533.

~~29~~30. Hertkorn N, Harir M, Koch BP, Michalke B, & Schmitt-Kopplin P (2013) High-field NMR spectroscopy and FTICR mass spectrometry: powerful discovery tools for the molecular level characterization of marine dissolved organic matter. Biogeosciences 10(3):1583-1624.

~~30~~31. Liu ZF, Sleighter RL, Zhong JY, & Hatcher PG (2011) The chemical changes of DOM from black waters to coastal marine waters by HPLC combined with ultrahigh resolution mass spectrometry. Estuar Coast Shelf S 92(2):205-216.

~~31~~32. Stubbins A, et al. (2010) Illuminated darkness: molecular signatures of Congo River dissolved organic matter and its photochemical alteration as revealed by ultrahigh precision mass spectrometry. Limnol. Oceanogr. 55:1467-1477.

~~32~~33. Stenson AC, Marshall AG, & Cooper WT (2003) Exact Masses and Chemical Formulas of Individual Suwannee River Fulvic Acids from Ultrahigh Resolution Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectra. Anal. Chem. 75(6):1275-1284.

~~33~~34. Dittmar T (2008) The molecular level determination of black carbon in marine dissolved organic matter. Org. Geochem. 39(Copyright (C) 2011 American Chemical Society (ACS). All Rights Reserved.):396-407.

~~34~~35. Boutegrabet L, et al. (2012) Attachment of Chloride Anion to Sugars: Mechanistic Investigation and Discovery of a New Dopant for Efficient Sugar Ionization/Detection in Mass Spectrometers. Chemistry – A European Journal 18(41):13059-13067.

~~35~~36. Flerus R, Koch BP, Schmitt-Kopplin P, Witt M, & Kattner G (2011) Molecular level investigation of reactions between dissolved organic matter and extraction solvents using FT-ICR MS. Marine Chemistry 124(1–4):100-107.

~~36~~37. Centre national pour l'exploitation des océans (France) (1983) Manuel des analyses chimiques en milieu marin (CNEXO, Brest) p 395.

~~37~~38. Gaspar A, et al. (2009) Combined utilization of ion mobility and ultra-high-resolution mass spectrometry to identify multiply charged constituents in natural organic matter. Rapid Commun. Mass Spectrom. 23(5):683-688.

~~38~~39. Stenson AC, Landing WM, Marshall AG, & Cooper WT (2002) Ionization and Fragmentation of Humic Substances in Electrospray Ionization Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry. Anal. Chem. 74(17):4397-4409.

~~39~~40. Hertkorn N, et al. (2008) Natural Organic Matter and the Event Horizon of Mass Spectrometry. Analytical Chemistry 80(23):8908-8919.

~~40~~

41. van Krevelen DW (1950) Graphical-statistical method for the study of structure and reaction processes of coal. Fuel 29:269-284.

42. Kendrick E (1963) A mass scale based on CH2=14.0000 for high resolution mass spectrometry of organic compounds. Anal. Chem. 35:2146.

43. Shakeri Yekta S, Gonsior M, Schmitt-Kopplin P, & Svensson BH (2012) Characterization of Dissolved Organic Matter in Full Scale Continuous Stirred Tank Biogas Reactors Using Ultrahigh Resolution Mass Spectrometry: A Qualitative Overview. Environ. Sci. Technol. 46(22):12711-12719.

44. Stedmon CA, Markager S, & Bro R (2003) Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. Marine Chemistry 82(3–4):239-254.

~~41~~45. Murphy KR, Stedmon CA, Graeber D, & Bro R (2013) Fluorescence spectroscopy and multi-way techniques. PARAFAC. Analytical Methods 5(23):6557-6566.

~~42~~46. Sleighter RL, Liu Z, Xue J, & Hatcher PG (2010) Multivariate Statistical Approaches for the Characterization of Dissolved Organic Matter Analyzed by Ultrahigh Resolution Mass Spectrometry. Environ. Sci. Technol. 44(19):7576-7582.

~~43~~47. Werner E, et al. (2008) Mass spectrometry-based metabolomics: Accelerating the characterization of discriminating signals by combining statistical correlations and ultrahigh resolution. Analytical chemistry 80(13):4918-4932.

~~44~~48. Stubbins A, et al. (2014) What's in an EEM? Molecular Signatures Associated with Dissolved Organic Fluorescence in Boreal Canada. Environ. Sci. Technol. 48(18):10598-10606.

~~45~~49. Chen M, Price RM, Yamashita Y, & Jaffe R (2010) Comparative study of dissolved organic matter from groundwater and surface water in the Florida coastal Everglades using multi-dimensional spectrofluorometry combined with multivariate statistics. Applied Geochemistry 25(6):872-880.

~~46~~50. Bro R & Smilde AK (2014) Principal component analysis. Analytical Methods 6(9):2812-2831.

~~47~~51. Gonsior M, et al. (2009) Photochemically Induced Changes in Dissolved Organic Matter Identified by Ultrahigh Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. Environ. Sci. Technol. 43:698-703.

~~48~~52. Kuhnert N, Drynan JW, Obuchowicz J, Clifford MN, & Witt M (2010) Mass spectrometric characterization of black tea thearubigins leading to an oxidative cascade hypothesis for thearubigin formation. Rapid Commun. Mass Spectrom. 24(23):3387-3404.

~~49~~53. Galindo C & Del Nero M (2014) Molecular Level Description of the Sorptive Fractionation of a Fulvic Acid on Aluminum Oxide Using Electrospray Ionization Fourier Transform Mass Spectrometry. Environ. Sci. Technol. 48(13):7401-7408.

~~50~~54. Aucour AM, et al. (2003) The Amazon River: behaviour of metals (Fe, Al, Mn) and dissolved organic matter in the initial mixing at the Rio Negro/Solimões confluence. Chem Geol 197(1–4):271-285.

~~51~~55. Wagner S, et al. (2015) Linking the Molecular Signature of Heteroatomic Dissolved Organic Matter to Watershed Characteristics in World Rivers. Environ. Sci. Technol.

~~52~~56. Yamaguchi, Y., Y. Matsubara, T. Ochi, T. Wakamiya, and Z.-I. Yoshida. 2008. How the π Conjugation Length Affects the Fluorescence Emission Efficiency. Journal of the American Chemical Society 130: 13867-13869.

57. Sharpless CM & Blough NV (2014) The importance of charge-transfer interactions in determining chromophoric dissolved organic matter (CDOM) optical and photochemical properties. Environmental Science: Processes & Impacts 16(4):654-671.

~~53. Gonsior M, et al. (2014) Changes in Dissolved Organic Matter during the Treatment Processes of a Drinking Water Plant in Sweden and Formation of Previously Unknown Disinfection Byproducts. Environ. Sci. Technol. 48(21):12714-12722.~~

~~54~~58. McClain ME, Richey JE, & Reynaldo LV (1995) Andean Contributions to the Biogeochemistry of the Amazon River System. Bull. Inst. fr. études andines 24(3):425-437.

~~55. Kim S, Kramer RW, & Hatcher PG (2003) Graphical Method for Analysis of Ultrahigh-Resolution Broadband Mass Spectra of Natural Organic Matter, the Van Krevelen Diagram. Anal. Chem. 75(20):5336-5344.~~

~~56~~59. Zech, W. and others 1997. Factors controlling humification and mineralization of soil organic matter in the tropics. Geoderma 79: 117-161.

~~. van Krevelen DW (1950) Graphical-statistical method for the study of structure and reaction processes of coal. Fuel 29:269-284.~~

~~57. Kendrick E (1963) A mass scale based on CH2=14.0000 for high resolution mass spectrometry of organic compounds. Anal. Chem. 35:2146.~~

~~58. Shakeri Yekta S, Gonsior M, Schmitt-Kopplin P, & Svensson BH (2012) Characterization of Dissolved Organic Matter in Full Scale Continuous Stirred Tank Biogas Reactors Using Ultrahigh Resolution Mass Spectrometry: A Qualitative Overview. Environ. Sci. Technol. 46(22):12711-12719.~~

Table 1: Differences in SPE-DOM between the Rio Negro, Rio Tapajós and Rio Madeira identified by ESI-FT-ICR-MS.

Intensity-weighted average values of all assigned molecular formulae (180-800 Da)								
Formulae	sample	n^*	Center of mass	$\Delta O:C_w$	$\Delta H:C_w$	ΔDBE_w	ΔDBE_w-O	$\Delta DBE_w:C$
CHO	Negro	3130	453	0.51	1.05	11.15	0.31	0.524
	Tapajós	3298	428	0.49	1.12	10.04	0.13	0.494
	Madeira	3151	419	0.51	1.10	9.93	-0.03	0.504
CHNO	Negro	1362	429	0.49	0.99	11.58	1.85	0.590
	Tapajós	1975	413	0.49	1.02	10.92	1.68	0.580
	Madeira	2088	408	0.50	1.01	10.91	1.64	0.592
CHOS	Negro	58	313	0.22	1.84	2.49	-0.99	0.143
	Tapajós	286	315	0.30	1.66	3.65	-0.75	0.237
	Madeira	399	357	0.49	1.21	7.15	-0.23	0.460

Note: $\Delta O:C_w$: Intensity-weighted averaged oxygen to carbon ratios of assigned molecular formulae; $\Delta H:C_w$: Intensity-weighted averaged hydrogen to carbon ratios of assigned molecular formulae; ΔDBE_w: Intensity-weighted averaged double bond equivalency; ΔDBE_w-O: ΔDBE_w with numbers of oxygen atoms subtracted, $\Delta DBE_w:C$: carbon normalized ΔDBE_w. n: number of assigned molecular formulae to *m/z* molecular ions.

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: Font: 10 pt, Font color: Black, Border: : (No border)

Formatted: Font: 10 pt

Table 2: Averaged Nutrients, chlorophyll, DOC and TDN concentrations of all water samples collected in the main stem of the Rio Negro, Rio Madeira, Rio Jamari and Rio Tapajós as well as in 8-10 flooded lakes within the Rio Negro area, Rio Madeira area as well as within the Rio Tapajós and after its confluence with the Amazon River (Tapajós area) in May 2013.

	DOC_{av} (mg/L)	DOC range (mg/L)	TDN (mg/L)	TDP (mg/L)	Chlorophyll a (µg/L)
Rio ~~Tapajós~~Negro area (10 flooded lakes and river)	~~3.9~~10.1	~~2.9-~~8.6-11.5~~.3~~	0.~~10~~12	0.~~03~~01	3
Rio ~~Negro~~Tapajós area (9 flooded lakes and river)	~~10.1~~3.9	~~8.6-11.~~2.9-5.3	0.~~12~~10	0.~~01~~03	3
Rio Madeira area (10 flooded lakes and river)	2.9	1.9-5.8	0.11	0.03	2
Rio Tapajós (~~main stem~~mainstem river only)	3.6		0.13	0.05	2
Rio Negro (~~main stem~~mainstem river only)	10.6		0.12	0.02	n.d
Rio Madeira (~~main stem~~mainstem river only)	5.8		0.18	0.07	n.d
Rio Jamari (~~main stem~~mainstem river only) (white water tributary to the Madeira River)	2.4		0.23	0.04	2

Note: DOC_{av}: averaged DOC concentrations from each area. Standard error of DOC measurements were always between 0.1 and 0.2 mg~~/L~~ $^{-1}$. TDN: Total dissolved nitrogen; TDP: Total dissolved phosphorous. n.d. = not detected. Standard errors of TDN and TDP wer always between 0.001 and 0.0025 mg $^{-1}$

Formatted: Font: 10 pt, Font color: Black, Border: : (No border)

Formatted: Inget, Font: Calibri, 11 pt

Formatted: Inget, Font: Calibri, 11 pt

Formatted: Inget, Font: Calibri, 11 pt

Formatted: Inget, Font: Calibri, 11 pt

Formatted: Inget, Font: Calibri, 11 pt

Formatted: Inget, Font: Calibri, 11 pt

Formatted: Justified

Formatted: Font color: Black, Superscript, Border: : (No border)

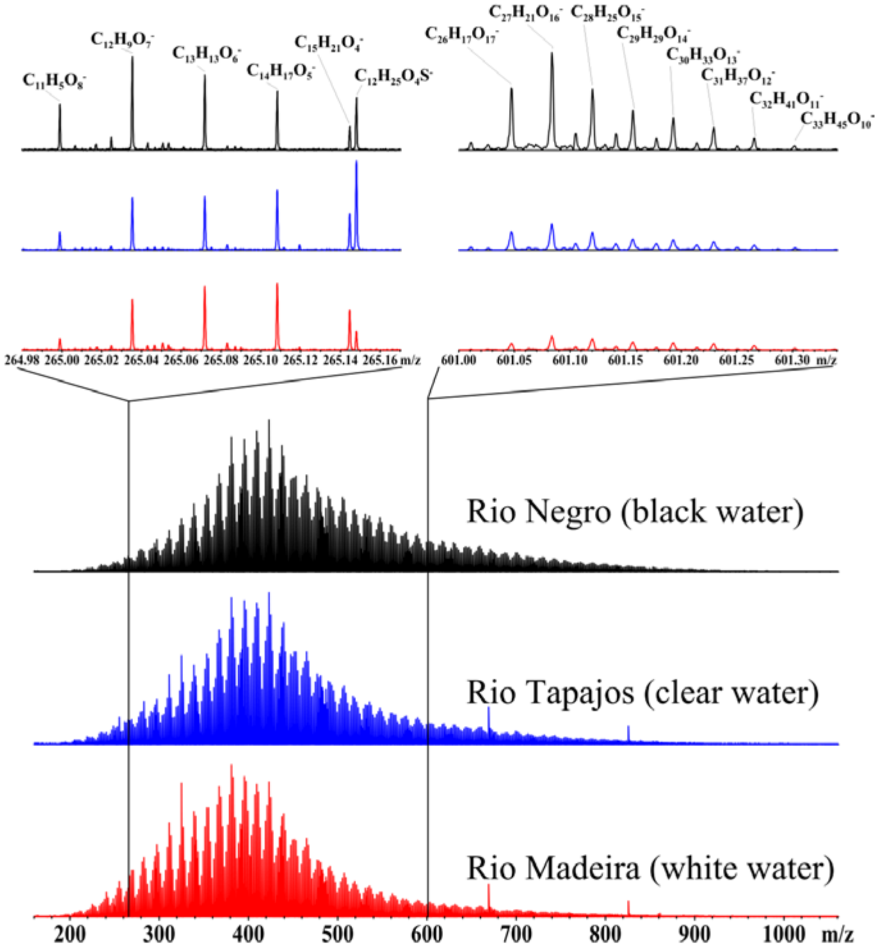


Figure 2: Ultrahigh resolution mass spectra of SPE-DOM isolated from the mainstem of the Rio Negro, Rio Madeira and Rio Tapajós, Amazon, Brazil and the stacked relative abundances of all ions at nominal mass 265 and 601 for all three river systems (note the diferences in peak height for different colors, slightly phase shifted for increased visibility).

Formatted: Font: 10 pt

Formatted: Font: 10 pt

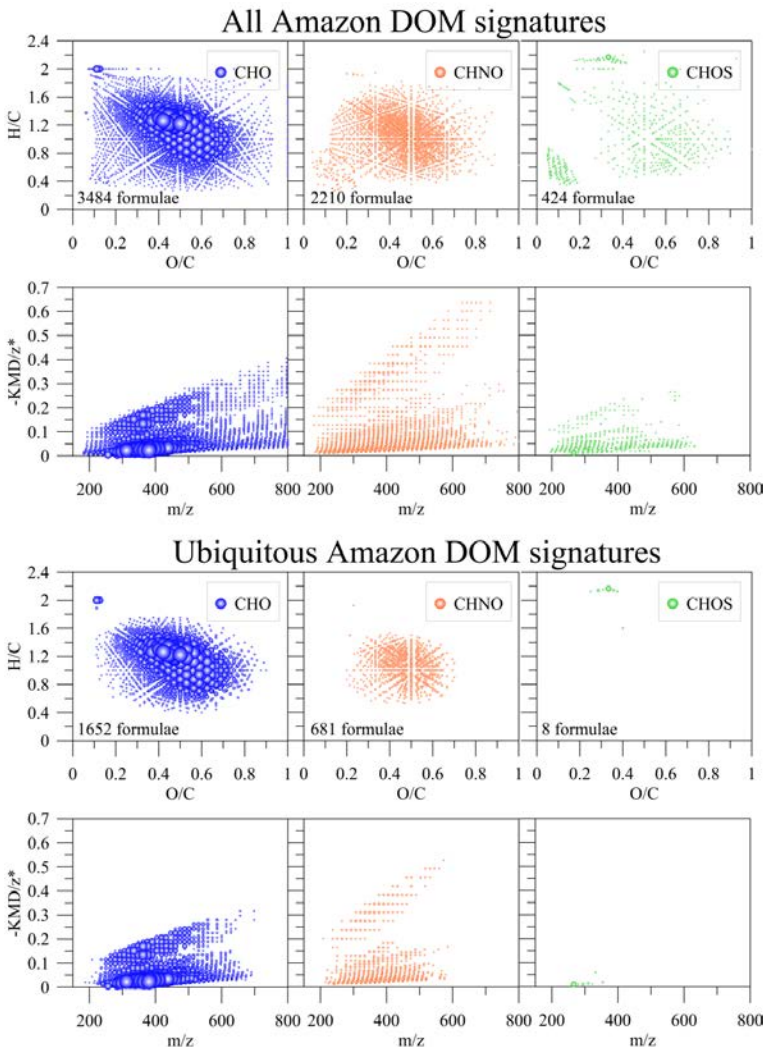


Figure 2: Van Krevelen diagrams of the total and the ubiquitous molecular ions in all Amazon SPE-DOM samples analyzed by FT-ICR-MS.

Formatted: Font: 10 pt

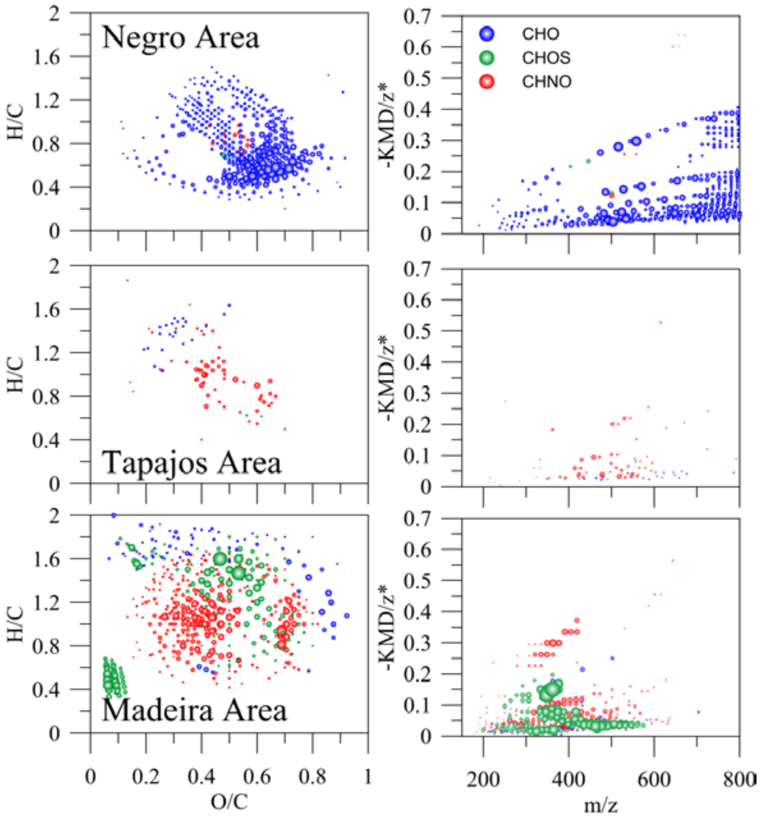


Figure 3: Unique ions analyzed by FT-ICR-MS of solid-phase extracted DOM associated with the three different Amazon areas.

Formatted: Font: 10 pt

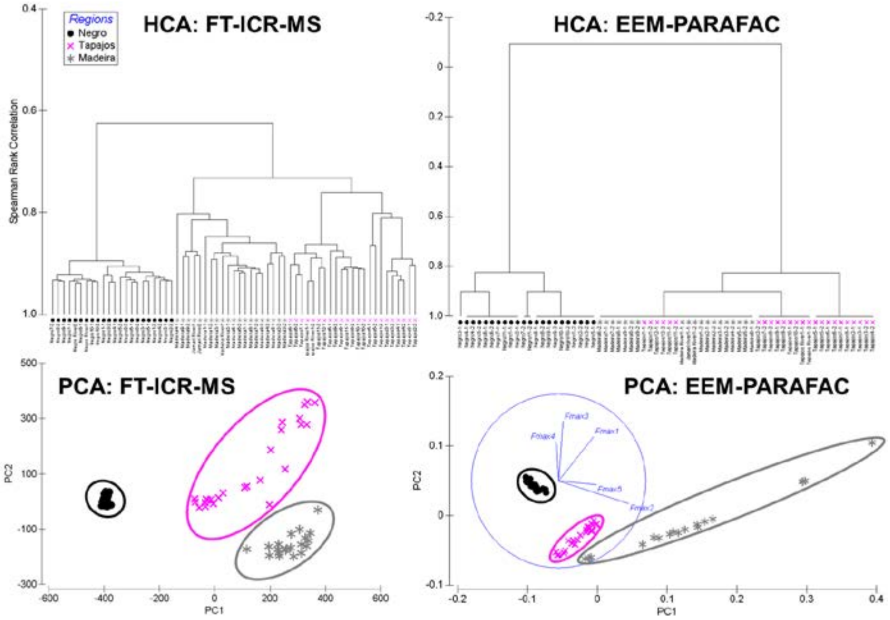
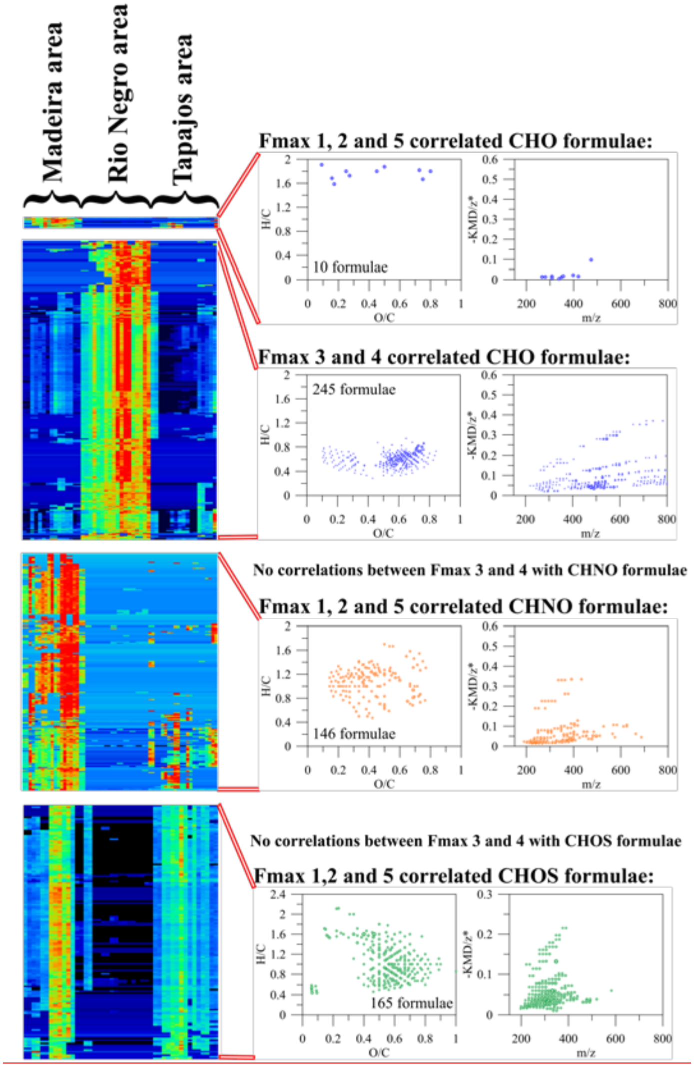


Figure 4: Principal component analysis and hierarchical cluster analysis of the normalized data of all mass peaks and their intensities and all EEM-PARAFAC Fmax components.

Formatted: Font: 10 pt


Madeira area
Rio Negro area
Tapajos area
Fmax 1, 2 and 5 correlated CHO formulae:
H/C
O/C
-KMD/z*
m/z
10 formulae
Fmax 3 and 4 correlated CHO formulae:
245 formulae
No correlations between Fmax 3 and 4 with CHNO formulae
Fmax 1, 2 and 5 correlated CHNO formulae:
146 formulae
No correlations between Fmax 3 and 4 with CHOS formulae
Fmax 1,2 and 5 correlated CHOS formulae:
165 formulae
0
0.2
0.4
0.6
0.8
1
1.2
1.6
2
2.4
0.1
0.3
0.5
200
400
600
800

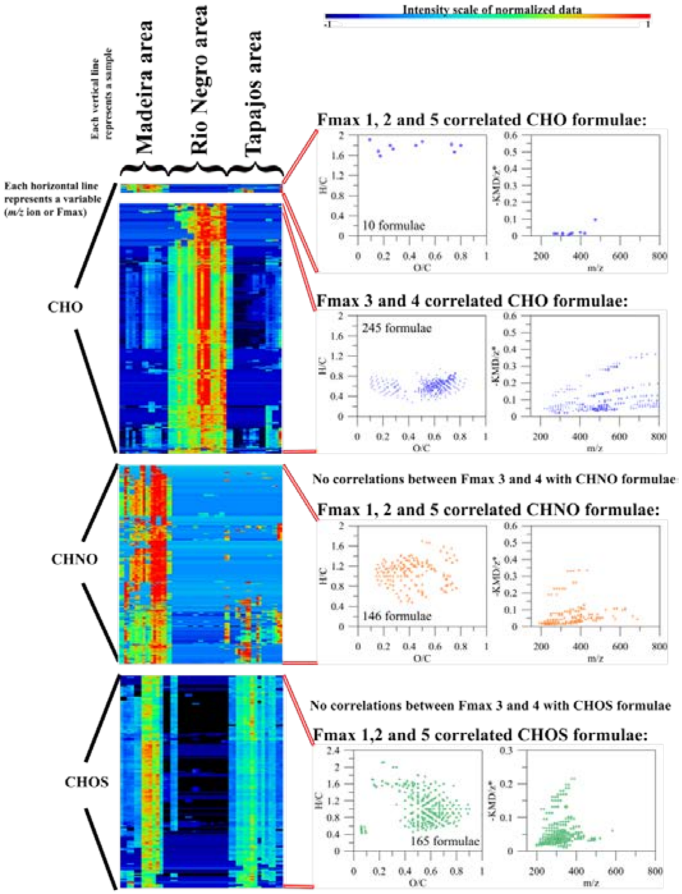


Figure 5: Heat maps of specific Spearman Rank correlated hierarchical clusters (distance threshold 0.6) that showed correlations between ~~molecular~~*m/z* ions and ~~their intensities and~~ Fmax values separated into the dissolved organic carbon (CHO) ~~and~~ nitrogen (CHNO) ~~pools. No correlations were found in the~~ and sulfur (CHOS) pool. ~~This approach was based on hierarchical clusters that contained the Fmax values to emphasize the tight relationships between Fmax values and the specific DOM pools.~~ The van Krevelen diagrams correspond to the molecular signatures that co-varied with specific EEM-PARAFAC Fmax values. The ~~whole hierarchical cluster~~entire heat maps and ~~other~~additional specific *m/z* ~~hierarchical~~ion clusters ~~with m/z intensities that co-varied~~ are given in Figure S6, including their distribution in the chemical space visualized in van Krevelen diagrams.

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: Font: 10 pt, Italic

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: Font: 10 pt