



Chemodiversity of Dissolved Organic Matter in the Amazon Basin

Michael Gonsior¹, Juliana Valle², Philippe Schmitt-Kopplin^{3,4}, Norbert Hertkorn⁴, David Bastviken⁵, Jenna Luek¹, Mourad Harir⁴, Wanderley Bastos⁶, and Alex Enrich-Prast^{2,5}

¹Chesapeake Biological Laboratory, University of Maryland Center for Environmental Science, Solomons, MD20688, USA

5 ²Departamento de Ecologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 21941-901, Brazil

³Analytical Food Chemistry, Technische Universität München, Freising-Weihenstephan, 85354, Germany

⁴Research Unit Analytical BioGeoChemistry, Helmholtz Zentrum Muenchen, Neuherberg, 85764, Germany

⁵Department of Thematic Studies - Environmental Change, Linköping University, Linköping, 581 83, Sweden

⁶Laboratory of Environmental Biogeochemistry, Universidade Federal do Rondônia, Rodovia, 76801-974, Brazil

10 *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-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 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. 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.

15
20
25

1 Introduction

With an average of about 200,000 m³/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

30



fate of particulate and dissolved organic matter throughout the Amazon River and its tributaries using a variety of methodologies. The sources of riverine OM have been identified through isotopic analysis of $\delta^{13}\text{C}$, $\delta^{18}\text{O}$, and $\delta^{15}\text{N}$ (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 $\delta^{14}\text{C}$ 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 have traditionally been classified as “black waters”, “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 of OM, such as through mediating light availability for aquatic life or photoreactivity. Black waters are influenced by chromophoric DOM (CDOM) and has 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 $\sim 2/3$ 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, these waters are not as light limited 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 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.

High bacteria counts ($\sim 10^9 \text{ L}^{-1}$) 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). Younger, biologically available organic matter is preferentially respired upstream and the apparent $\delta^{14}\text{C}$ 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 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). 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 these techniques, we attempted to answer a) What is the overall chemodiversity of DOM in the Amazon Basin, and moreover, do distinct differences in DOM composition 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 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

Madeira, Negro and Tapajós River surface water samples were collected in duplicate by 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, 8-10 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 collected samples were either filtered and solid phase extracted directly after collection (aboard river boats) or within three hours after collection, when smaller boats were used for sampling.

2.1 Solid-Phase Extraction

All 1 L water samples were filtered through pre-combusted 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), but it was modified by using formic acid instead of hydrochloric acid (HCl). The reason behind this choice of acid, was that the remaining chloride ions in HCl may cause chloride ion adduct formation in the ionization source of the FT-ICR-MS (34). Briefly, SPE-cartridges (Agilent, Bond Elut PPL, 1g resin), were activated with methanol (Chromasolv, Sigma Aldrich), rinsed with acidified ultrapure water (Milli-Q) and the acidified water samples were gravity-fed through the cartridges at a flow rate of ~10 mL/min.

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



5 glass vials. Samples were kept cool and later frozen at $-18\text{ }^{\circ}\text{C}$. Methanolic extracts are stable when kept at $-18\text{ }^{\circ}\text{C}$ for extended periods of time (35). 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 samples varied between 3-10 mg L⁻¹ and hence only 1 L of each sample was used to make sure that the used 1 g SPE cartridges were not overloaded.

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 and phosphorus analyses were performed by an automated colorimetric flow injection analysis (FIA) system, manufactured by FOSS®, according to the quality control guidelines recommended by FOSS. 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).

2.3 Ultrahigh Resolution Mass Spectrometry

15 A mass accuracy of less than 0.2 ppm deviation from the actual mass is often 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): however, in our analysis we only encountered singly charged ions as described previously (38). 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).

20 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.

25 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.

2.3 Excitation Emission Matrix Fluorescence and Parallel Factor Analysis

30 The used SPE method recovered CDOM almost quantitatively (>90%) and in order to directly compare FT-ICR-MS results and optical properties of Amazon DOM, we also used SPE-DOM for the determination of optical properties. Hundred μL of methanolic SPE-DOM sample 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. 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) 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). Several models were tested and a 5 component model was developed on the normalized data (Fig. S3). 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). 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). 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 used for all subsequent statistical analyses.

15 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) and EEM-PARAFAC (45). 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). 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.

25 In an additional analysis, the two normalized data sets were combined to be able to determine hierarchical clusters on the variables (m/z ions and their intensities and EEM-PARAFAC Fmax values). This approach enabled the generation of heat maps (software: TM4-Multi Experiment Viewer) and clusters of m/z ions and also which m/z ions co-varied with Fmax values. This approach successfully depicted clusters of specific classes of m/z ions with distinct and confined chemodiversity as expressed in van Krevelen diagrams within the CHO, CHNO and CHOS pools (Fig. S6), and revealed subsets of m/z ions that correlated well with EEM-PARAFAC components.



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. 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 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 were observed for the molecular ions that represented nitrogen-containing compounds, but the numbers of assigned CHNO formulae were noticeably 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 Amazon Basin 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.

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.

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 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 subunits 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).

A removal of these unique and likely disk-shaped high molecular weight polyaromatic DOM molecules by means of mineral adsorption or flocculation with iron (21) or aluminum oxides (49) after mixing with the high sediment-load Solimões River is conceivable and corresponded to previous reports that between 4% (50) and 40% (2) of the DOC was removed at the confluence of the Rio Negro and Solimões. Investigations of Amazon wetland hydrogeochemistry 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 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 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 anthropogenic influence as suggested in a recent study (51) and this suggestion was supported by the relatively high population close 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 to grow soy beans.

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 in all investigated areas.

EEM spectra of SPE-DOM also showed distinct differences between each area (Fig. S2), and a high intensity long wavelength fluorescent peak in Rio Negro water samples was apparent. A five PARAFAC components (Fmax1-5) model (Fig. S3) was most adequate to explain the differences in the fluorescence data set and also captured the indicative fluorescence signal of the Rio Negro (Fmax4) at high emission wavelengths. These high emission wavelengths can either be explained by large complex conjugated π -systems that would support the unique aromatic high molecular weight DOM characteristic of Rio Negro waters or charge transfer processes (52), 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 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, presumably non-light absorbing aliphatic CHOS and CHNO compounds were indicative for the Madeira sampling area and might have been responsible for some of the smaller differences seen in the HCA between the FT-ICR-MS and EEM-PARAFAC data (Fig. 4). Aromatic CHNO compounds, present in the Madeira area samples, might 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 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), and the suggested adsorption or more likely coagulation of DOC when mixed with high suspended sediment rivers. High Rio Negro DOC concentrations may be indicative of watershed characteristics (17) or a lower mineral content (2, 7, 50) compared to the other tributaries. The DOC data were in agreement with the observations resulting from FT-ICR-MS and EEM-PARAFAC analyses 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).

To address the question 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 unique molecular ions found in the Rio Negro area (Fig. 3), hydrogen-deficient (low H/C ratios), but highly oxygenated (high O/C ratio) molecular ions were clustered together



including a strong positive correlation with all Rio Negro samples (Fig. S6). Similarly, the Fmax components 3 and 4 were highly correlated with an analogously aromatic cluster of molecular ions with high molecular weight (Fig. 5 and S6). 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 Fmax 3 and 4 did not show any correlations with molecular ions in the CHNO pool supporting the suggestion 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 Fmax 1,2 and 5 correlated always together and so 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).

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. 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 and therefore it is more likely that this 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. 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, 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 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. 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.

10 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.
- 15 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.
- 20 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.
- 25 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.
- 5 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
10 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
15 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-
20 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.
- 25 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. Gonsior M, et al. (2011) Characterization of dissolved organic matter across the Subtropical Convergence off the
30 South Island, New Zealand. *Marine Chemistry* 123(1-4):99-110.
25. 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. 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. Tremblay LB, Dittmar T, Marshall AG, Cooper WJ, & Cooper WT (2007) Molecular characterization of dissolved
5 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. Ward ND, et al. (2013) Degradation of terrestrially derived macromolecules in the Amazon River. *Nature Geosci* 6(7):530-533.
29. Hertkorn N, Harir M, Koch BP, Michalke B, & Schmitt-Kopplin P (2013) High-field NMR spectroscopy and
10 FTICR mass spectrometry: powerful discovery tools for the molecular level characterization of marine dissolved organic matter. *Biogeosciences* 10(3):1583-1624.
30. 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. Stubbins A, et al. (2010) Illuminated darkness: molecular signatures of Congo River dissolved organic matter and
15 its photochemical alteration as revealed by ultrahigh precision mass spectrometry. *Limnol. Oceanogr.* 55:1467-1477.
32. 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. Dittmar T (2008) The molecular level determination of black carbon in marine dissolved organic matter. *Org.*
20 *Geochem.* 39(Copyright (C) 2011 American Chemical Society (ACS). All Rights Reserved.):396-407.
34. 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. Flerus R, Koch BP, Schmitt-Kopplin P, Witt M, & Kattner G (2011) Molecular level investigation of reactions
25 between dissolved organic matter and extraction solvents using FT-ICR MS. *Marine Chemistry* 124(1–4):100-107.
36. Centre national pour l'exploitation des océans (France) (1983) Manuel des analyses chimiques en milieu marin (CNEXO, Brest) p 395.
37. 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.
- 30 38. 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. Hertkorn N, et al. (2008) Natural Organic Matter and the Event Horizon of Mass Spectrometry. *Analytical Chemistry* 80(23):8908-8919.



40. 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. Murphy KR, Stedmon CA, Graeber D, & Bro R (2013) Fluorescence spectroscopy and multi-way techniques. *PARAFAC. Analytical Methods* 5(23):6557-6566.
- 5 42. 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. 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. Stubbins A, et al. (2014) What's in an EEM? Molecular Signatures Associated with Dissolved Organic
10 Fluorescence in Boreal Canada. *Environ. Sci. Technol.* 48(18):10598-10606.
45. 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. Bro R & Smilde AK (2014) Principal component analysis. *Analytical Methods* 6(9):2812-2831.
- 15 47. 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. 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.
- 20 49. 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. 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.
- 25 51. Wagner S, et al. (2015) Linking the Molecular Signature of Heteroatomic Dissolved Organic Matter to Watershed Characteristics in World Rivers. *Environ. Sci. Technol.*
52. 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.
- 30 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. 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. 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 $CH_2=14.0000$ 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.

10

**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.

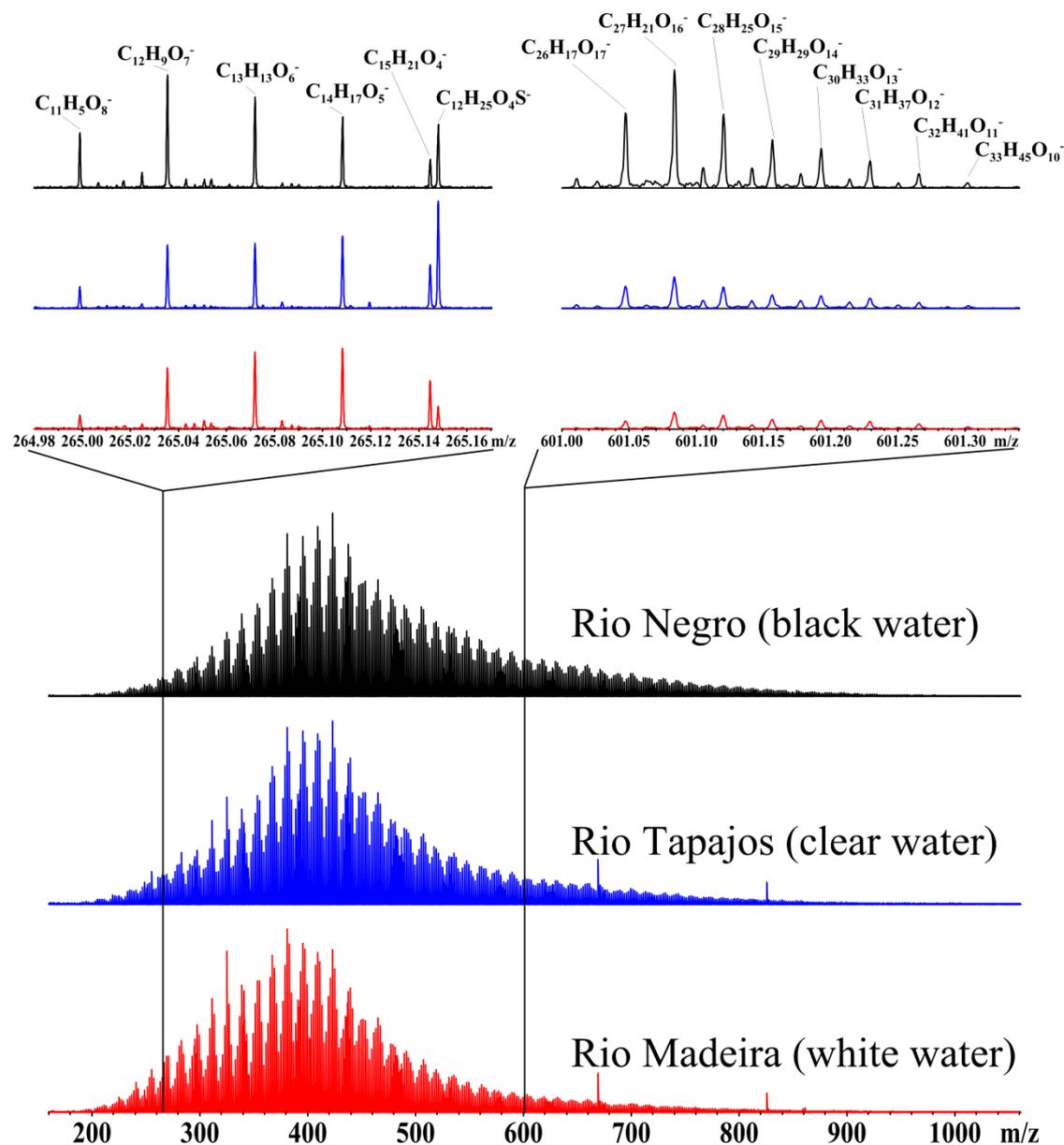


Table 2: Averaged Nutrients, chlorophyll, DOC and TDN concentrations of all water samples collected in the main stem of the Rio Negro, Rio Madeira, Rio Jamari and Rio Tapajós as well as in 8-10 flooded lakes within the Rio Negro area, Rio Madeira area as well as within the Rio Tapajós and after its confluence 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 area (flooded lakes)	3.9	2.9-5.3	0.10	0.03	3
Rio Negro area (flooded lakes)	10.1	8.6-11.5	0.12	0.01	3
Rio Madeira area (flooded lakes)	2.9	1.9-5.8	0.11	0.03	2
Rio Tapajós (main stem river)	3.6		0.13	0.05	2
Rio Negro (main stem river)	10.6		0.12	0.02	n.d
Rio Madeira (main stem river)	5.8		0.18	0.07	n.d
Rio Jamari (main stem 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. TDN: Total dissolved nitrogen; TDP: Total dissolved phosphorous. n.d. = not detected

5



5 Figure 2: Ultrahigh resolution mass spectra of SPE-DOM isolated from 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 differences in peak height for different colors, slightly phase shifted for increased visibility).

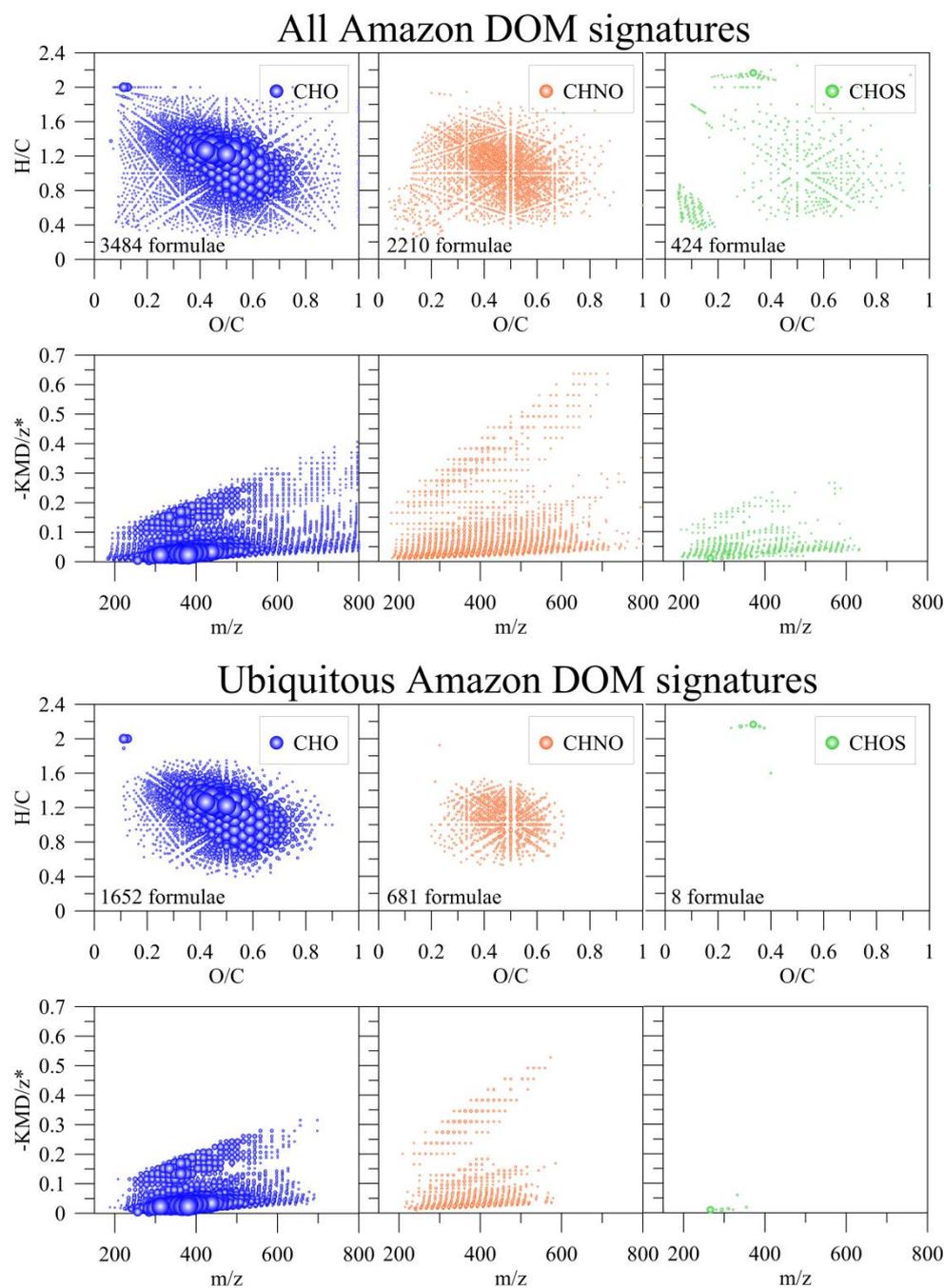


Figure 2: Van Krevelen diagrams of the total and the ubiquitous molecular ions in all Amazon SPE-DOM samples analyzed by FT-ICR-MS.

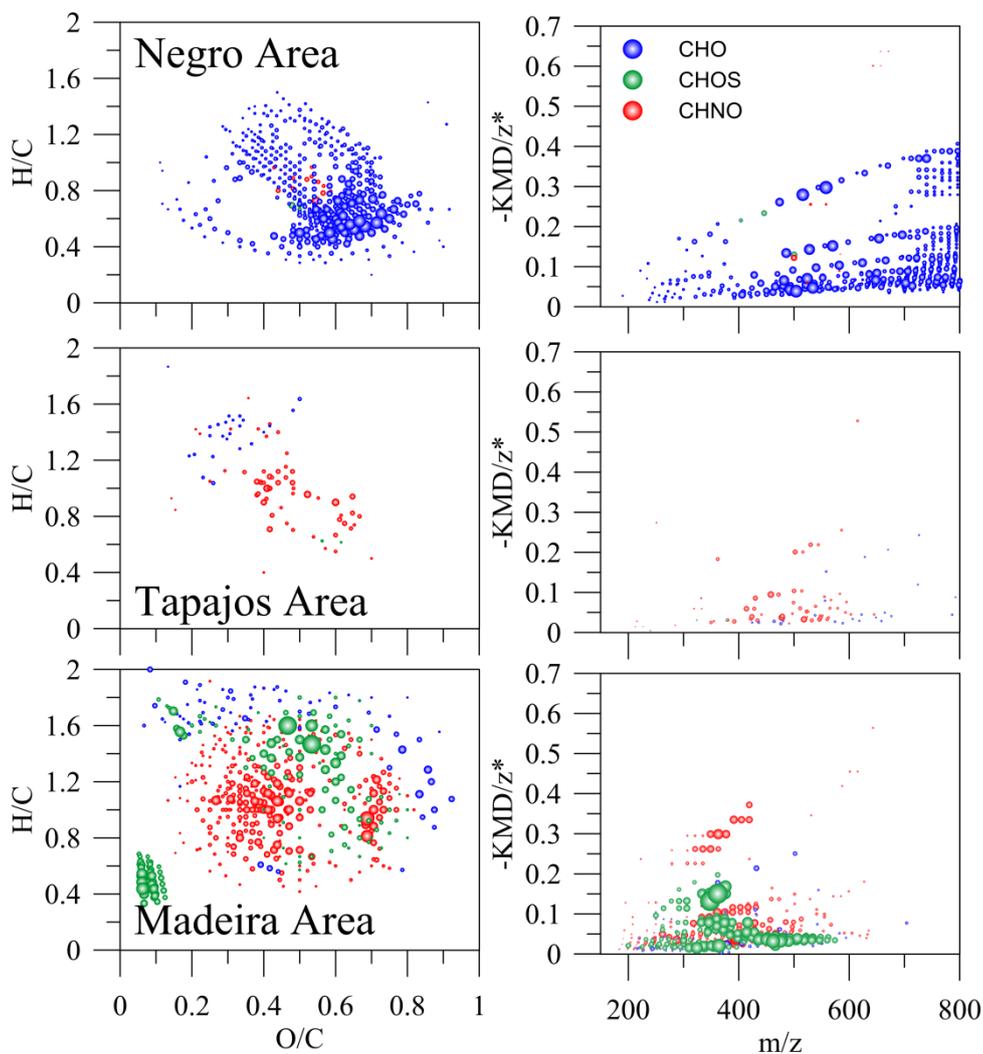


Figure 3: Unique ions analyzed by FT-ICR-MS of solid-phase extracted DOM associated with the three different Amazon areas.

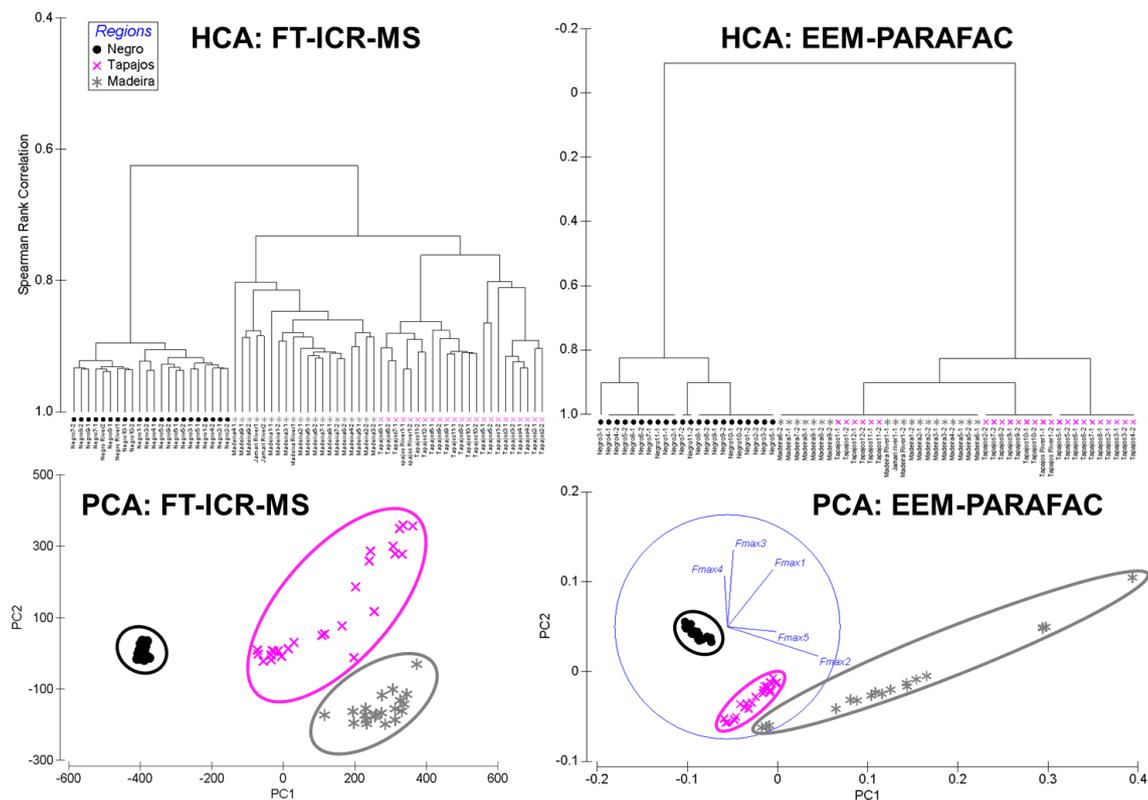
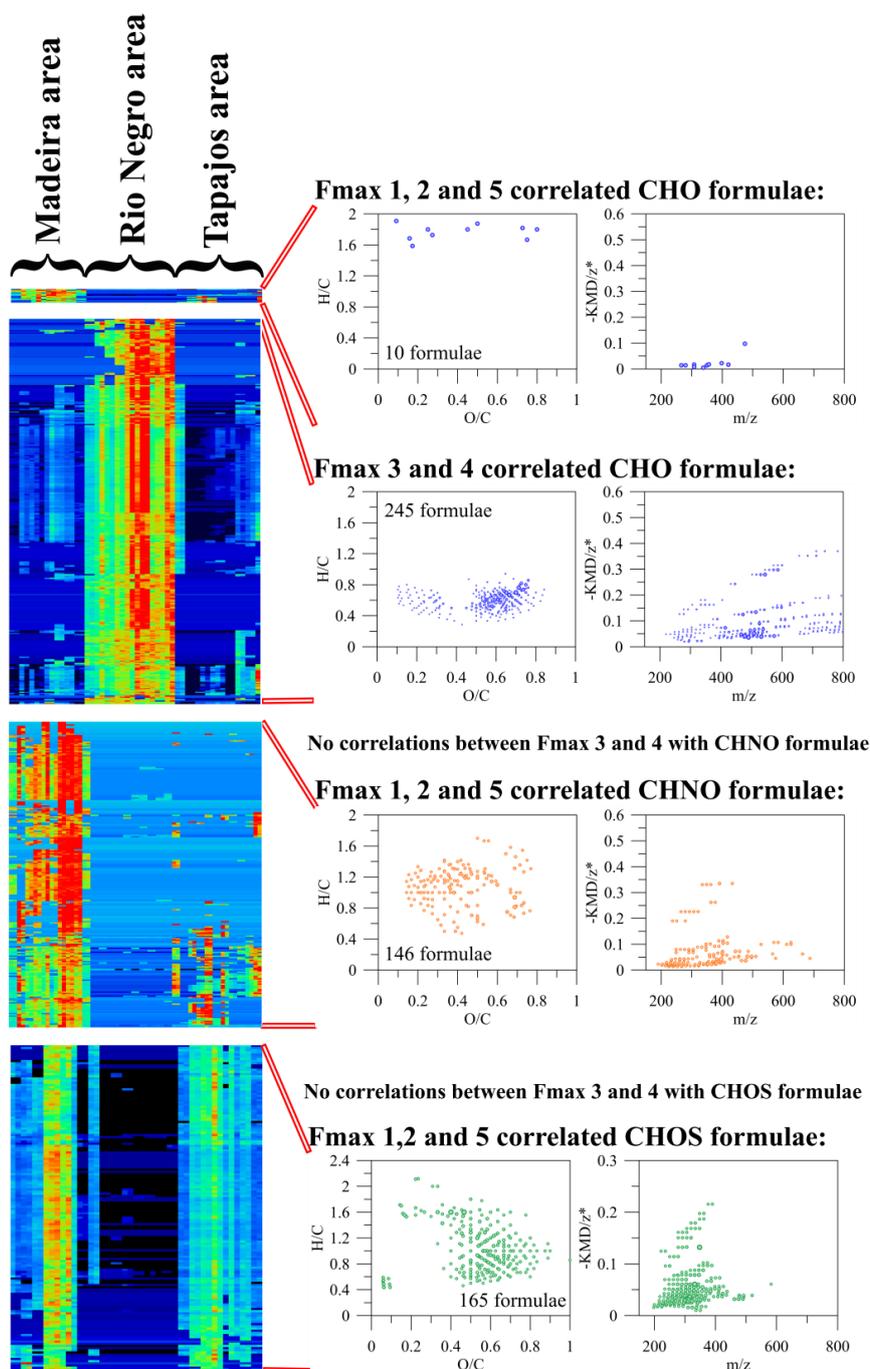


Figure 4: Principal component analysis and hierarchical cluster analysis of the normalized data of all mass peaks and their intensities and all EEM-PARAFAC Fmax components.

5



5 **Figure 5: Heat maps of Spearman Rank correlations between molecular ions and their intensities and Fmax values separated into the dissolved organic carbon (CHO) and nitrogen (CHNO) pools. No correlations were found in the 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 and other m/z intensities that co-varied are given in Figure S6, including their distribution in the chemical space visualized in van Krevelen diagrams.**