

Metabolomics analysis of unresolved molecular variability in stoichiometry dynamics of a stream dissolved organic matter

Bahareh Hassanpour^a, Neal Blair^{a,b}, Ludmilla Aristilde^{a,*}

^a Department of Civil and Environmental Engineering, McCormick School of Engineering and Applied Science, Northwestern University, Evanston, IL 60208, United States

^b Department of Earth and Planetary Sciences, Weinberg College of Arts and Sciences, Northwestern University, Evanston, IL 60208, United States

ARTICLE INFO

Keywords:

Carbon transport
Dissolved organic matter
Mass spectrometry
Phytochemicals
Phenylpropanoids
Fatty acids

ABSTRACT

Broad molecular classification based on stoichiometric ratio relationships has been used extensively to characterize the chemical diversity of aquatic dissolved organic matter (DOM). However, variability in the molecular composition within this classification has remained elusive, thus limiting the interpretation of DOM dynamics, especially with respect to transport versus transformation patterns in response to hydrologic or landscape changes. Here, leveraging high-frequency spatiotemporal sampling during rainfall events at a Critical Zone Observatory project site in Clear Creek, Iowa, we apply a metabolomics-based analysis validated with fragmentation using tandem mass spectrometry to uncover patterns in the molecular features of the DOM composition that were not resolved by classification based on stoichiometric ratios in the chemical formulae. From upstream to downstream sites, beyond the increased aromaticity implied by changes in the stoichiometric ratios, we identified an increased abundance of flavonoids and other phenylpropanoids, two important subgroups of aromatic compounds. The stoichiometric analysis also proposed a localized decline in the abundance of lipid-like compounds, which we attributed specifically to medium-chain and short-chain fatty acids; other lipids such as long-chain fatty acids and sterol lipids remained unchanged. We further determined in-stream molecular transitions and specific compound degradation by capturing changes in the molecular masses of terpenoids, phenylpropanoids, fatty acids, and amino acids. In sum, the metabolomics analysis of the chemical formulae resolved molecular variability imprinted on the stoichiometric DOM composition to implicate key molecular subgroups underlying carbon transport and cycling dynamics in the stream.

1. Introduction

The composition of stream dissolved organic matter (DOM), which is considered to be of terrestrial (allochthonous) sources from surrounding landscapes or of aquatic (autochthonous) sources due to in-stream metabolic activities, can be influenced by biological, chemical, and physical processes (Bauer and Bianchi, 2012; Fasching et al., 2020, 2016; Giorgio and Pace, 2021; Lynch et al., 2019; Stutter et al., 2013; Tanentzap et al., 2019). Therefore, along a stream path, the diversity of DOM composition is expected to depend on the source and levels of DOM in the headwaters, the fresh DOM input from the landscape, and the extent of DOM transformation within the stream (Creed et al., 2015; Giorgio and Pace, 2021; Kamjunke et al., 2019; Lynch et al., 2019; Singh et al., 2014). Traditionally, the chemical composition and functional properties of natural DOM have been inferred using chromophoric and fluorescent properties (Bhattacharya and Osburn, 2020; Vione et al.,

2021; Yang et al., 2015). Recent applications of high-resolution mass spectrometry (HRMS)-based approaches using primarily Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) have led to an unprecedented molecular investigation of DOM dynamics in terrestrial waters based on the stoichiometric ratio representation of chemical formulae often illustrated by a Van Krevelen diagram (Gonsior et al., 2011; Maizel et al., 2017; McDonough et al., 2020; Phungsai et al., 2016; Sleighter and Hatcher, 2008; Valle et al., 2020; Zhang et al., 2018). To date, Van Krevelen analyses have provided valuable information on DOM composition, dynamics, and transformations in aquatic environments (Pang et al., 2021; Riedel et al., 2016). However, the Van Krevelen diagram used as a classical two-dimensional stoichiometry-dependent chemical class assignment, primarily based on three elements [carbon (C), oxygen (O), and hydrogen (H)], provides a limited categorization of biomolecules and thus may lead to incorrect classification of molecular classes (Rivas-Ubach et al., 2018). A recently

* Corresponding author.

E-mail address: ludmilla.aristilde@northwestern.edu (L. Aristilde).

<https://doi.org/10.1016/j.watres.2022.118923>

Received 4 May 2022; Received in revised form 20 July 2022; Accepted 28 July 2022

Available online 29 July 2022

0043-1354/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

proposed multidimensional stoichiometric compound (MSC) classification approach (Rivas-Ubach et al., 2018), which was validated by including six abundant elements [C, H, O, nitrogen (N), phosphorus (P), and sulfur (S)], enabled the classification of compounds into five groups (phytochemicals, carbohydrates, nucleotides, proteins, lipids, and amino sugars)—these molecular groups refer to broad classes of biomolecules, except that “phytochemicals” represent a diverse group of keto-aromatic biomolecules that could be derived from plants, algae, and microorganisms. Despite the diverse molecular profiles of compounds within each stoichiometry-based molecular grouping, within-group molecular variability has not been explored. Such high-resolution molecular information is necessary for the accurate evaluation of the stream DOM dynamics and transformation (Garayburu-caruso et al., 2020; Lynch et al., 2019), which is crucial in determining lability, persistence, metabolism, and functional identity of the DOM composition (Cyle et al., 2020; Lusk and Toor, 2016; Lynch et al., 2019; Remucal et al., 2012).

Water flow promotes the connectivity of carbon pools between the landscape and aquatic environments and contributes to the flushing of organic matter from riparian areas and soil surfaces into surrounding streams (Fasching et al., 2016; Lynch et al., 2019; Singh et al., 2014).

The DOM export during storm events, which can comprise two-thirds of the stream DOM load annually, can introduce rapid changes in the molecular features in the DOM composition along the stream path (Creed et al., 2015; Oeurng et al., 2011). This DOM export can result in elevated levels of vascular plant-derived organic matter with high molecular complexity and aromaticity from terrestrial sources whereas base flow-associated DOM may contain a high fraction of protein-like and lipid-like compounds derived from autochthonous sources (Lu and Liu, 2019). Moreover, increased heteroatom content of the DOM observed in the downstream areas (including wetlands) are often attributed to autochthonous sources due to microbial and algal metabolism (Fasching et al., 2020; Fellman et al., 2009). Beyond broad classification based on chromophoric and fluorescent properties and stoichiometric analysis, still lacking is an in-depth and high-resolution characterization to capture molecular characteristics of spatial and temporal variations of stream DOM composition.

Here, we seek to obtain this molecular characterization for the stream DOM in Clear Creek, Iowa, a site extensively characterized by the Intensively Managed Landscapes – Critical Zone Observatory (IML-CZO) project and other studies (Papanicolaou et al., 2010; Hou et al., 2018; Kim et al., 2020; Blair et al., 2021 and citations within). High-resolution

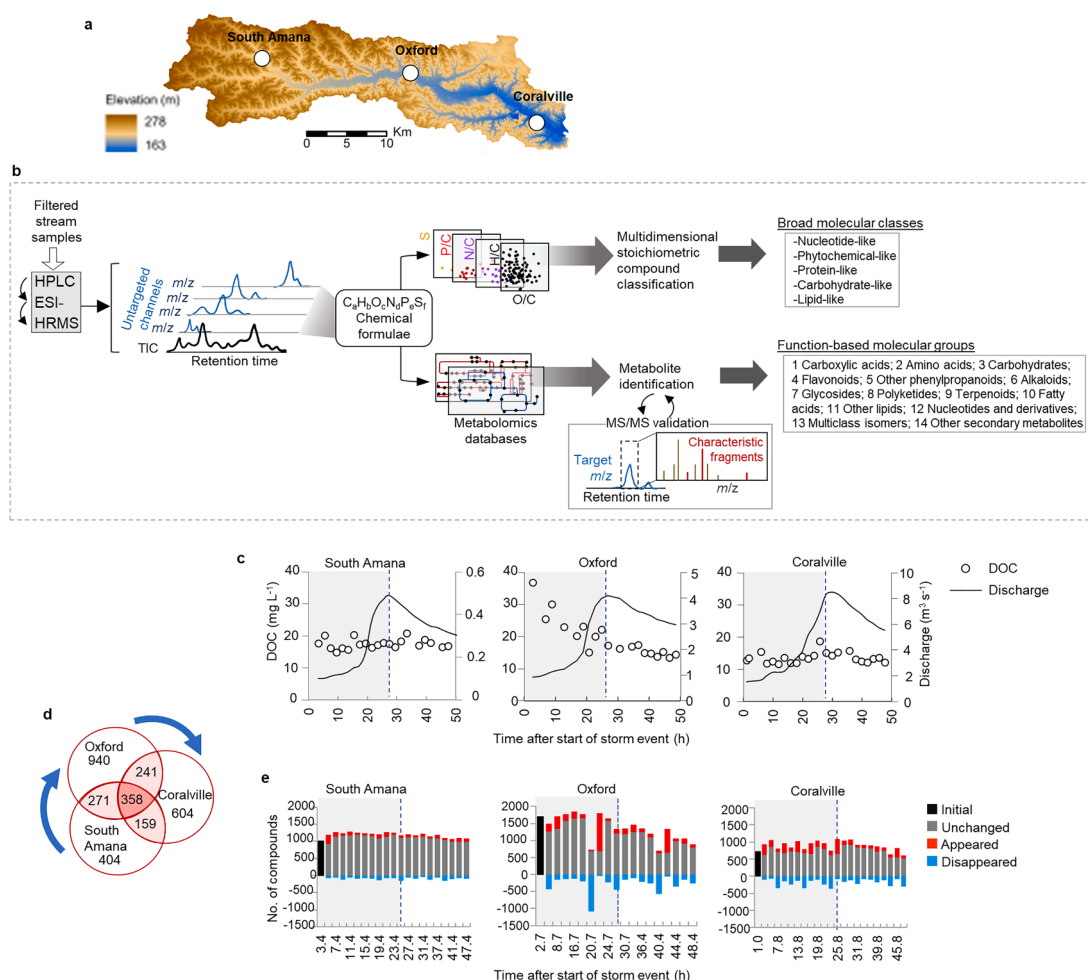


Fig. 1. Overview of study site and analysis workflow, and DOC concentrations and compound abundance at the three sampling locations over the storm duration. **a**, The watershed topography and the sampling locations, South Amana (upstream), Oxford (midstream), and Coralville (downstream). **b**, Analysis workflow of compound classification using stoichiometric ratios versus metabolomics-based analysis. **c**, The DOC concentrations (mg L⁻¹) and discharge (m³ s⁻¹) at each sampling point during the storm event. **d**, Similarities and differences between the number of unique chemical formulae at each site. The values in overlapping areas in the interlocked circles indicate the number of common chemical formulae. **e**, variation in the number of unchanged, appeared, and disappeared compounds in comparison to the previous sampling point at the three sites. Note the different scale of the secondary y-axis in **c** due to nearly a difference of two orders of magnitude for the range of discharge values at the three sites. Abbreviations in **b**: HPLC, high-performance liquid chromatography; ESI, electrospray ionization operated in negative mode; HRMS, high-resolution mass spectrometry. The gray areas in **c** and **e** indicate the rising limb of the hydrograph.

time-resolved stream sampling was performed at three locations along an approximately 45 km-stretch of Clear Creek immediately prior to and following a spring storm event (Fig. 1a; Supplementary Fig. S1). The upland agricultural portions of the watershed have steeper hillslopes that facilitate connectivity between the erodible surfaces and the channel (Fig. 1a). Runoff from floodplains and forested riparian buffers, which are prevalent downstream, intercept agricultural runoff from upstream (Blair et al., 2021). Therefore, the different landscape features of this watershed provide an appropriate case study to investigate the spatiotemporal dynamics in our molecular investigation of the stream DOM composition (Fig. 1a). The application of a metabolomics-based high-resolution chemical fingerprinting of small compounds (or metabolites) derived from plants and microorganisms is an attractive method to advance a molecular understanding of the source and fate of stream DOM (Lynch et al., 2019) (Fig. 1b). Using high-performance liquid chromatography (LC) coupled with orbitrap-based HRMS for an untargeted investigation, we characterized molecular variabilities in the stream DOM composition dynamics using metabolomics-based molecular identities and, for comparative analysis, stoichiometry-dependent MSC classification (Fig. 1b).

2. Site description and methods

2.1. Site description

Clear Creek, a tributary of the Iowa River (in east-central Iowa, USA), drains an approximate 275 km² watershed (Blair et al., 2021; Hou et al., 2018; Kim et al., 2020). The baseflow in this watershed is driven by groundwater and subsurface agricultural drainage (or tile drainage), which could contribute up to 50% of the streamflow (Schilling et al., 2012; Schilling and Helters, 2010). Soils in the Clear Creek watershed are silty clay to silt loam Mollisols and Alfisols, with an organic carbon content ranging from ~2 wt% in agricultural soils to ~4 wt% in restored prairie soils (Hou et al., 2018). Prior to the start of the growing season, which begins in early May and continues until early October, flash floods can result due to intense storms, lack of vegetation cover, and fine soil textures (Hou et al., 2018; Kim et al., 2020). The South Amana site captures a predominantly headwater signature from mostly tile-drained croplands dominated by corn-soybean crop rotations (Blair et al., 2021). Land use in the South Amana sub-basin comprised 86% agricultural, 6% developed, and 8% natural lands. The subtle V-shaped cross-section of the valley facilitates the transport of hillslope erosion from the fields to the channel. By contrast, the valley cross-section of the downstream sites, Oxford and Coralville, is more U- or box-shaped due to the presence of floodplains. The Oxford site is downstream of a small town (Oxford, Iowa) and receives water from a mixture of forest and grass lands (23%), though corn and soybean cultivation still dominates (69%) (Papanicolaou et al., 2010). The Coralville station integrates water from 92% of the Clear Creek watershed, comprising 60% agricultural and 28% natural lands. Urban land use, which is low (11%), is localized in the downstream portion of the Coralville sub-basin. The three sampling sites along Clear Creek capture DOM export from the upper sub-basin (South Amana), middle sub-basin (Oxford), and lower sub-basin (Coralville) locations (Fig. 1a). The drainage area for these three sites is 26.2, 157.5, and 254.3 km², respectively (Blair et al., 2021). The geomorphology of this river corridor has been implicated in the temporal (within a storm event) and spatial (downstream transport) responses of particulate organic carbon (POC) in previous studies (Blair et al., 2021; Kim et al., 2020). Water discharge at South Amana was determined using a stage-discharge relationship, as previously detailed (Blair et al., 2021). Discharge for Oxford and Coralville were obtained from the USGS website (<https://waterdata.usgs.gov/ia/nwis/>; gaging stations # 05454200 and # 05454300, respectively).

2.2. Sample collection

A total of 63 stream water samples were acquired from the three sampling locations at 2- to 4 h intervals over a near 48 h period to cover shortly before and during a spring storm event (March 29th to March 31st, 2017) in coordination with the IML-CZO (Kim et al., 2020). For each sample, ~800 mL of stream water was collected using ISCO auto-samplers (Teledyne ISCO) with 1 L Nalgene bottles after triple-rinsing the sampling line with stream water. After collection, the samples were filtered through pre-combusted fiber filters (0.7 µm pore size, Millipore) and were kept frozen until further analysis. The hourly precipitation data for nearby Iowa City (from the Iowa State Environmental Mesonet website, <https://mesonet.agron.iastate.edu/>), showed that a total of 32.5 mm of rainfall occurred during the study period.

2.3. Dissolved organic carbon (DOC) and metabolomics-based analyses

First, using an Apollo 9000 TOC Combustion Analyzer (Teledyne Tekmar Co.), we quantified the DOC concentration at each sampling point for each station. For this TOC analysis, standards were run at the start, in the middle, and at the end of the analysis run for each set of samples analyzed for independent quality checks. Second, we obtained mass spectra of the DOM in the stream samples using a LC-HRMS approach. After filtration (0.2 µm nylon) of thawed samples, two 1 mL aliquots of each sample were concentrated 20 times by drying under a gentle flow of nitrogen gas (99.9% purity) before dissolving in ultra-pure water (Thermo Scientific™ Barnstead™) followed by centrifugation for 10 min at 10,000 g and 4 °C. The supernatants were analyzed by high-performance LC equipped with a ion-paired reverse phase column (Thermo Scientific Dionex UltiMate 3000) with HRMS (Thermo Scientific Q Exactive Hybrid Quadrupole-Orbitrap), which ran in full-scan (72–840 m/z) negative ion mode (Aristilde et al., 2017b; Kukurugya et al., 2019). The MS instrument was calibrated weekly to an accuracy of < 0.5 ppm. For LC separations, we used a 25-min gradient run (solvent A: 3% methanol/15 mM acetic acid/10 mM tributylamine and Solvent B: 100% methanol) with a Waters Acquity UPLC BEH C18 (1.7 µm × 2.1 mm × 100 mm), at the flow rate of 180 µL min⁻¹. The solvent gradient with respect to solvent A was the following: 0 min, 100%; 2.5 min, 100%; 5 min, 80%; 7.5 min, 80%; 10 min, 45%; 12 min, 45%; 14 min, 5%; 17 min, 5%; 18 min, 0%; 25 min, 0%. All solvents were Optima™ LC/MS Grade (Fisher Chemical™).

Focusing on compounds in the DOM composition with < 840 Da, compound peak identification and subsequent grouping of spectral features were conducted using the XCMS package (v 3.14.0) in R programming Language (v 4.1.0) (Benton et al., 2010; Smith et al., 2006; Tautenhahn et al., 2008). Signals were obtained with a 25-ppm mass tolerance and a minimum intensity of 10⁴. The adducts and natural isotopes were found using the CAMERA package in R (Kuhl et al., 2012). Following deduction of chloride ion adducts and isotopic daughter ions, the spectral features were further refined by applying blank subtraction (min signal/blank ratio=5) and identifying peaks in technical duplicates using an in-house code in R programming environment (Venables et al., 2020). To annotate chemical formulae, the OmicsCraft online database (<http://tools.omicscraft.com/MetaboQuest/>) was used to search the m/z against different compound databases (PubChem, HMDB, LIPID MAPS, KEGG, and MMCD) with the mass tolerance of 20 ppm. The majority of annotated chemical formulae (75%) had a mass error of ≤ ±10 ppm. First, the annotated chemical formulae were categorized using the stoichiometry-dependent MSC classification approach, which considers the heteroatom content of biomolecules (C, H, O, N, P, S) and four different stoichiometric ratios (O/C, H/C, N/C, P/C) to classify compounds as carbohydrate-like (carbohydrates and amino sugars), lipid-like, nucleotide-like, phytochemicals, or protein-like (Rivas-Ubach et al., 2018). Second, the annotated chemical formulae were subjected to metabolomics-based classification by cross-referencing the returned results with KEGG Brite hierarchy files (<https://www.kegg.jp>) using an

R code followed by manually scrutinizing each group for accuracy. For simplicity of presentation, the metabolites were grouped into 14 categories: flavonoids, phenylpropanoids, terpenoids, alkaloids, polyketides, glycosides, fatty acids, other lipids (e.g., fatty acyls and sterol lipids), carbohydrates, amino acids (including dipeptides), carboxylic acids, nucleotides and their associated derivatives (nucleotides, nucleosides, and nucleobases), other secondary metabolites (betalains, natural toxins, vitamins, cofactors, hormones, and transmitters), and multiclass isomers (chemical formulae with multiple entries in the KEGG database regarding their functional categorization). We employed collision-induced MS/MS (or tandem MS) analysis to obtain fragmentation patterns to resolve uncertainties in determining chemical structures, especially when the metabolomics-based classification of a compound differed from or was not determined by the MSC classification. Data-dependent acquisition (top 5 peaks) with normalized collision energies of 15, 30, and 45 were used on a selected number of samples (4 from each site). Fragmentation spectra were acquired at resolution 17,500; precursor ions were isolated using 1 m/z width. Analysis of MS/MS fragmentation spectra was performed in R using XCMS package (Benton et al., 2010; Smith et al., 2006; Tautenhahn et al., 2008). Extracted fragments were analyzed using Xcalibur software (Thermo Scientific), and when available, the METLIN tandem mass spectrometry database was used for compound identification (Smith et al., 2005).

2.4. Statistical analysis

Statistical analysis (F-test, *t*-test, and multivariate analysis) and data visualization based on kernel density estimation were performed in JMP (Statistical Discovery™ from SAS Institute Inc.). Multivariate analysis including principal component analysis (PCA) and redundancy analysis is widely used to connect metrics of DOM composition and landscape characteristics (Bhattacharya and Osburn, 2020; Frost et al., 2006; Graeber et al., 2012). We conducted the PCA analysis to evaluate the correlation between the molecular groupings of chemical formulae with respect to the discharge and location along the stream. For this analysis, to account for the impact of increased flow from upstream to downstream, discharge (*Q*) was normalized by maximum discharge at each site (*Q*_{max}). In addition, the abundance for each molecular group at each site was normalized by the total number of characterized compounds at each site.

3. Results

3.1. Spatiotemporal DOM variations as a function of discharge

First, we examined changes in the DOC levels in response to the storm event at the three site locations (Fig. 1a, c). At both South Amana and Coralville, there was no clear trend between the discharge and the DOC concentrations (Supplementary Fig. S2), which varied from 14.9 mg L⁻¹ to 20.9 mg L⁻¹ and from 11.7 mg L⁻¹ to 18.7 mg L⁻¹, respectively (Fig. 1c). However, at Oxford, a steady decrease in DOC concentration from 36.8 mg L⁻¹ to 13.4 mg L⁻¹ during the storm event reflected a DOC dilution at this site, which would result from a limited DOC input as a function of the increased discharge at this site (Singh Dhillon and Inamdar, 2014) (Fig. 1c). The highest compound diversity was found at the midstream Oxford site (Fig. 1d). While we annotated 2441 chemical compounds (1810 unique chemical formulae, by excluding isomers) at Oxford, we only found a total of 1509 chemical compounds (1192 unique chemical formulae) at South Amana and 1633 chemical compounds (1361 unique chemical formulae) at Coralville (Fig. 1d). Markedly, the identification of a relatively small amount (358) of common chemical formulae across all three site locations highlighted the spatial variation in the DOM chemical composition in Clear Creek, implying a fast turnover of the DOM consistent with the lack of increasing DOC concentration during the storm (Fig. 1d).

At South Amana, there was a small temporal variability in the

number of compounds (from 1007 to 1089 from the start to the end of the study period), with 41% of the compounds consistently present at all samples across the sampling time (Fig. 1e; Supplementary Fig. S3). At Oxford, the dynamic appearance and disappearance of compounds at each subsequent sampling time resulted in an overall 50% decrease in the number of distinct compounds by the end of the sampling period (Fig. 1e). At Coralville, these dynamic changes led to an overall 15% decrease in the number of identified compounds by the end of the storm (Fig. 1e). Therefore, relatively greater temporal variations at Oxford and Coralville relative to South Amana during the storm implied the influence of different DOM source contributions or DOM transformations at the different sites (Supplementary Fig. S3). To obtain further insights into changes in the molecular identities of the compounds, we first evaluated changes in the DOM stoichiometry (Fig. 2).

3.2. DOM stoichiometry implies dynamics in molecular diversity

Most annotated compounds (> 80%) had O/C ratios < 0.5, which have been attributed previously to lipids, proteins, and lignin-derived compounds (Fig. 2a) (D'Andrilli et al., 2013; Lu et al., 2015; Rivas-Ubach et al., 2018). While the Oxford site had the most O-containing compounds consistent with its high compound diversity, the average O/C ratio (0.42) was the greatest at Coralville, highlighting the presence of relatively more oxidized compounds at the downstream location (Fig. 2a, Supplementary Fig. S5). At the beginning of the storm, compounds had well distributed H/C between 1 and 2.1 at all three sites, indicating the presence of diverse compounds in the stream before the rise of the discharge as was also reported for POC (Fig. 2a) (Rivas-Ubach et al., 2018). Approximately 20% of unsaturated compounds with high aromaticity and fewer alkyl groups (H/C < 1.3) disappeared by the peak of the flow, and those with H/C > 1.3 persisted at South Amana and at Oxford (Fig. 2a). During the receding limb of the hydrograph at Oxford, when the influence of subsurface interflow would be more pronounced than at the rising limb, compounds with relatively higher H/C ratios (1.5 < H/C < 2) became predominant (Fig. 2a). At the downstream location (Coralville), the average H/C ratio (=1.27) of the compounds was lower than the two upstream sites (average H/C = 1.37) (*p* < 0.05) (Fig. 2a; Supplementary Fig. S4), indicating a shift to unsaturated compounds both with fewer alkyl groups and with increased aromaticity downstream.

The proportion of N-containing chemical formulae (44–49%) was relatively higher than reported previously in terrestrial lakes (12–32%; Gonsior et al., 2011; Maizel et al., 2017), lake sediments (21–34%; Zhang et al., 2018), and the ocean (up to 10%; Sleighter and Hatcher, 2008) (Supplementary Fig. S5). At South Amana and Oxford, the N/C values of the compounds remained relatively constant (0 < N/C < 0.5) during the sampling period, but the peak discharge at Coralville revealed the transport of compounds with higher N/C values (Fig. 2a). Coralville had the highest average N/C value (0.26) among the three sites (*p* < 0.001; Supplementary Fig. S4). A small percentage (6–8%) of the compounds had P in their chemical composition (Fig. 2a) (Rivas-Ubach et al., 2018). The P/C ratio in most of these compounds remained the same (at about P/C < 0.1) but the density of P-containing compounds with this P/C ratio increased downstream over the sampling period (Fig. 2a). Most detected S-containing compounds had one S atom in their chemical composition at all three sites before and after the storm, stressing the persisting presence of mono-thiolated compounds such as cysteine and glutathione or the presence of sulfonated compounds widely detected in natural waters (Chu et al., 2016); low abundance of di-sulfated compounds (with two S), which could include cystine (a cysteine dimer), was also detected (Fig. 2a).

Instead of the widely-used Van Krevelen plots, whose compound classification is based on the atomic ratios of H/C versus O/C (Supplementary Fig. S6), we analyzed the abundance of compounds in relation to the atomic ratios of O/C, H/C, N/C, and P/C present in the chemical formulae using the MSC classification to obtain five stoichiometry-based

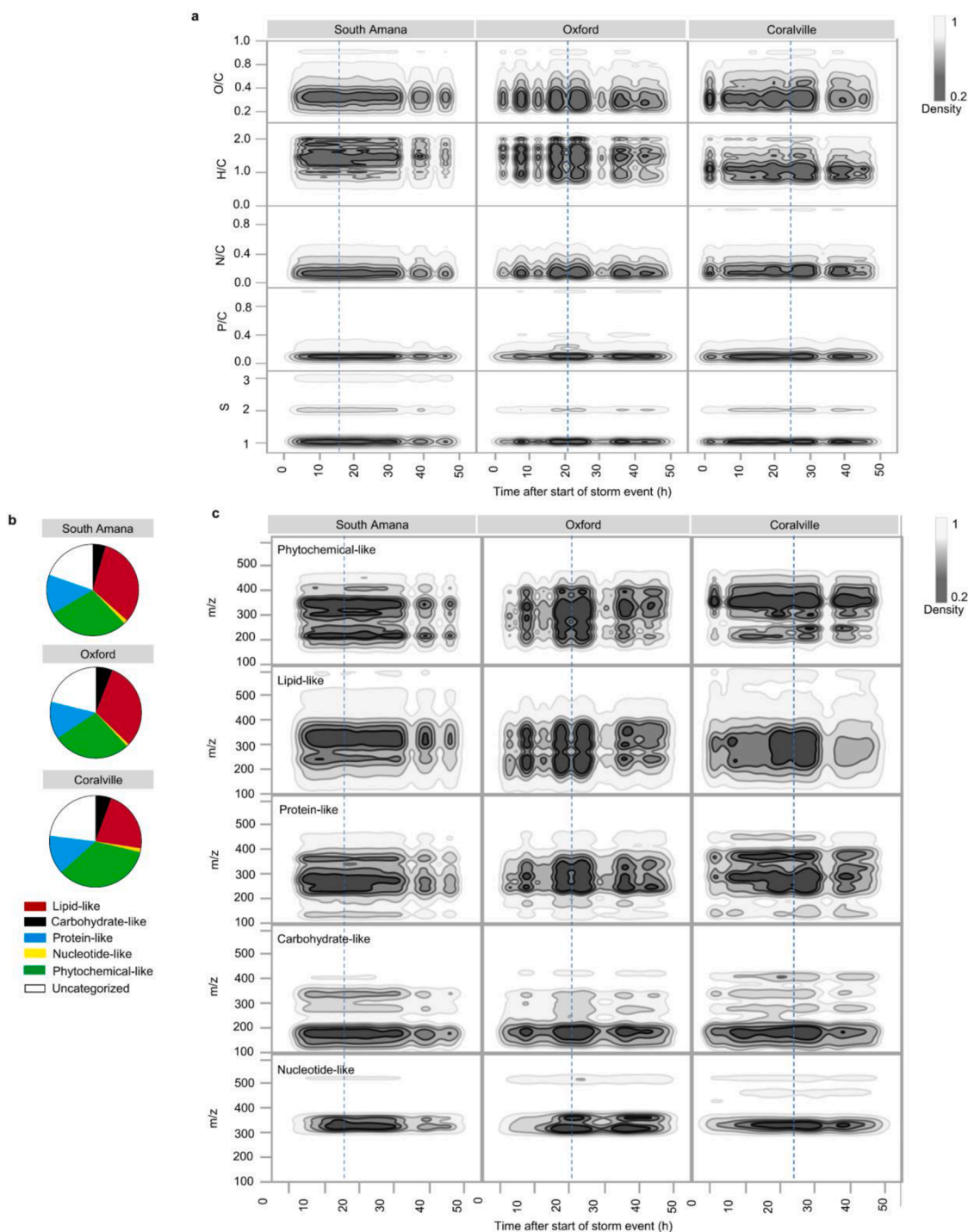


Fig. 2. Variability in spectral features with respect to stoichiometry and m/z during the storm event along the stream. **a**, Changes in atomic ratios of O/C, H/C, N/C, P/C, and the number of S atoms in chemical formulae during the storm event for the three sites. **b**, Proportion of all identified compounds categorized with respect to multidimensional stoichiometry compound (MSC) classification at each of three sites along Clear Creek. **c**, Changes in m/z values for compounds categorized by MSC classification approach. The illustrations in **a** and **c** are non-parametric density contours wherein the density number in the legend indicates the percentage of total compounds: white surfaces represent 100% of the compounds and dark gray surfaces represent 20% of the compounds.

groupings of the annotated compounds (Fig. 1b) (Rivas-Ubach et al., 2018). The majority of the compounds across the three sites were classified as phytochemical-like (28–34%) and lipid-like (22–31%) compounds, whereby the percentage of phytochemical-like compounds increased from upstream to downstream, accompanied by a decrease in the proportion of more labile lipid-like compounds (Fig. 2b and Supplemental Table S1), consistent with previous reports of preferential degradation of lipid-like compounds (Ladd et al., 2021). Protein-like

compounds comprised 12–13% of the annotated compounds (Fig. 2b). Carbohydrate-like (2–4%) and nucleotide-like (< 1%) compounds comprised a small fraction of the classified compounds (Fig. 2b). Regarding the phytochemical-like compounds, they were clustered primarily at two ranges of m/z of ~ 200 and at 350 at South Amana but, at Oxford, these compounds had a well-distributed m/z during the rising limb of the hydrograph, and, at Coralville, phytochemical-like compounds with lower m/z were less abundant (Fig. 2c). At the beginning of

the sampling period at all three sites, there were well-distributed m/z values (200–400) of lipid-like compounds (Fig. 2c). As discharge increased at South Amana, the abundance of low-molecular-mass lipid-like compounds ($m/z < 300$) diminished (Fig. 2c). By contrast, at Oxford, the lipid-like compounds with a wide range of m/z values declined steadily during the storm (Fig. 2c; Supplementary Fig. S7). And, at Coralville, the lipid-like compounds had relatively constant m/z values during the study period, but there was a surge in lipid-like compounds with $200 < m/z < 400$ after the early flush (~ first 30 h) (Fig. 2c). For protein-like compounds, the range of the m/z values at South Amana decreased at the crest of the hydrograph when the abundance of compounds with $m/z < 300$ declined but, at Coralville, these compounds appeared after the early flush, implying their upstream to downstream transport (Fig. 2c). The majority of carbohydrate-like and nucleotide-like compounds had a relatively constant abundance and m/z (150–200 m/z for carbohydrate-like compounds and ~ 350 m/z for nucleotide-like compounds) along the stream and during the study period (Fig. 2c; Supplementary Fig. S7). In sum, the stoichiometry-based data analysis implied, from upstream to downstream sites, an enrichment in aromatic plant-derived compounds and a preferential decomposition of lipid-like compounds, which were reported previously to be bioreactive (Ladd et al., 2021; Lu et al., 2015).

3.3. High-resolution metabolomics-based molecular characterization with fragmentation validation

To capture variability in the molecular-level composition of the aforementioned broad stoichiometry-based classification of the DOM dynamics, we subjected the LC-HRMS data to a metabolomics-based analysis scheme for the identification of specific compounds (or derivatives) from plant and microbial metabolism (Figs. 1b and 3). The compound identities based on the metabolomics-based analysis (13% of the total unique chemical formulae) were classified into the following 14 molecular groupings of metabolites with functional roles in metabolism: carboxylic acids, amino acids (including dipeptides), carbohydrates (including monosaccharides, disaccharides, and amino sugars), flavonoids, terpenoids, alkaloids, other phenylpropanoids, glycosides, polyketides, fatty acids, other lipids, nucleotides and their derivatives, multiclass isomers, and other secondary metabolites (Fig. 3). Towards validating our metabolomics-based molecular grouping, we also performed MS/MS fragmentation analysis to obtain molecular fragments determinants of specific molecular functional classes, particularly for compounds that were categorized differently by the MSC classification (Fig. 3a). For instance, tyrosine (an aromatic amino acid), which was considered a phytochemical-like compound by the broad MSC classification, was classified correctly as an amino acid by the metabolomics-based protocol through the identification of the carboxyl-amino moiety ($m/z = 72.0077$) (Fig. 3a). Guanosine, correctly binned as a nucleotide derivative through metabolomics-based assignment via the

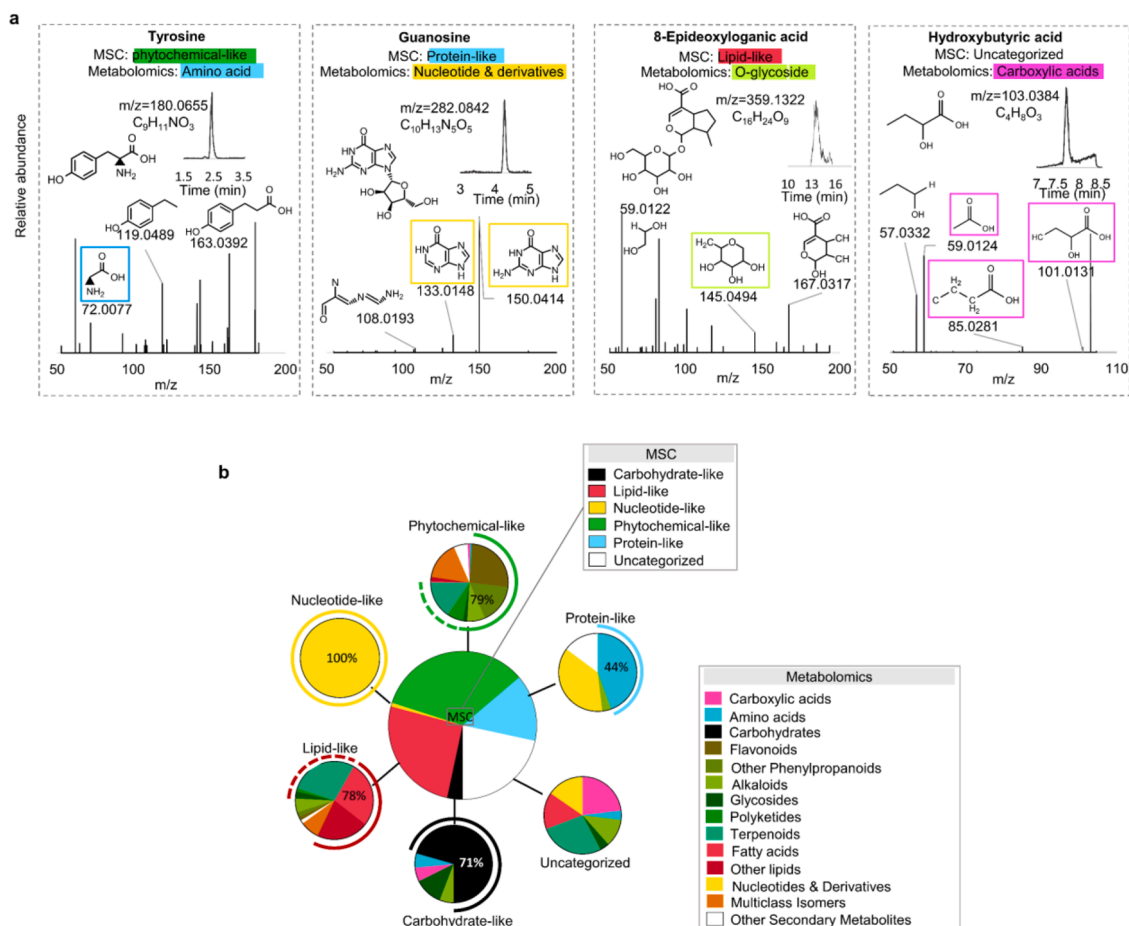


Fig. 3. Relationships and discrepancies between MSC classification and metabolomics-based classification of compounds. **a**, Examples of MS/MS fragmentation analysis of stream samples to confirm the predicted identity of compounds classified by the metabolomics-based approach versus the MSC classification approach. **b**, Compounds grouped using the metabolomics-based analysis versus being categorized by the broad MSC classification. In **b**, the inner large pie chart illustrates the distribution of the MSC-based categories; the outer small pie charts illustrate the diverse metabolomics-based compound groupings for the formulae categorized under each MSC classification; dashed lines around the outer pie charts indicate the compounds with multi-class metabolomics-based assignments.

guanine fragment ($m/z = 150.0414$), was classified incorrectly as a protein-like compound by MSC classification (Fig. 3a). Several glycosides, which generate a sugar fragment and at times contained lipid components, were categorized broadly as carbohydrate-like or lipid-like by the stoichiometry-based analysis (Fig. 3a and 3). For example, 8-epideoxyloganic acid is a terpene glycoside (or an iridoid glycoside), which contains both lipid and sugar moieties, but it was annotated as lipid-like compound by the MSC approach (Fig. 3a). Importantly, our metabolomics-based scheme identified carboxylic acids, which were among the uncategorized compounds or incorrectly annotated as carbohydrate-like compounds under the MSC classification (Fig. 3a and b).

For all the cross-categorized chemical formulae, the agreement between the MSC and metabolomics-based classifications varied between 44% and 100% (Fig. 3b). We found that 100% of the chemical formulae categorized as nucleotide-like compounds by the MSC classification were assigned as nucleotides (or nucleotide derivatives) by the metabolomics-assisted approach (Fig. 3b). Moreover, based on the metabolomics-based classification, 71% of the compounds classified as carbohydrate-like were found to be carbohydrates (Fig. 3b). However, among those compounds broadly grouped as protein-like by MSC classification, the metabolomics-based assignment determined that 44% of them belonged to amino acids (and dipeptides). Furthermore, due to stoichiometric similarities with small peptides as N-containing compounds, the MSC approach categorized nucleotides as protein-like compounds, but they were assigned accurately as nucleotides or nucleotide derivatives by the metabolomics analysis. We note, however, that proteins and large peptides were not characterized by the metabolomics-based approach (Fig. 3b). A large fraction (79%) of phytochemical-like compounds was assigned to compounds generally attributed to plant-derived compounds such as flavonoids, alkaloids and phenylpropanoids; a small fraction (5%) was annotated as other secondary metabolites such as betalains (Fig. 3b). Still, there are important groups of plant-derived compounds, including alkaloids and terpenoids, which were not categorized by the MSC approach but were captured by the metabolomics analysis (Fig. 3b). Furthermore, a diverse group of metabolomics-identified compounds including carboxylic acids (e.g., hydroxybutyric acid; Fig. 3a), short-chain fatty acids, nucleotides and their derivatives remained uncategorized by the MSC approach (Fig. 3b). In sum, these aforementioned differences in the compound categorizations by the metabolomics-based approach highlight the broad, mis-categorization, or lack of categorization by the MSC approach (Fig. 3b). Therefore, the metabolomics-based assignments offered additional and detailed molecular insights into the stream DOM dynamics beyond the molecular features proposed by the MSC classification (Fig. 3b). These new insights afforded a new picture of the molecular variability in the stream DOM as discussed in the following section.

3.4. Metabolomics-based analysis of spatial and temporal molecular variability in stream DOM

Here we present the insights into the spatiotemporal molecular transitions resulting from the metabolomics-based analysis. In relation to the dynamic changes in the broad molecular classes deduced from the stoichiometric patterns of the DOM, we analyzed the abundances of the metabolomics-based group of metabolites associated with each molecular class as well as within-group dynamics of the metabolites by monitoring changes in m/z along the stream (Fig. 4; Supplementary Figs. S7 and S8). At South Amana, during the rising limb of the hydrograph, we found that an initial increase (from 316 to 409) and subsequent decrease (to 362) in the abundance of lipid-like compounds coincided with increased abundance of fatty acids and other lipids at this site (Supplementary Figs. S7 and S8). At Oxford, among the decreased abundance observed for a diverse group of compounds during the storm, lipid-like compounds had the sharpest decline (from 577 to 282;

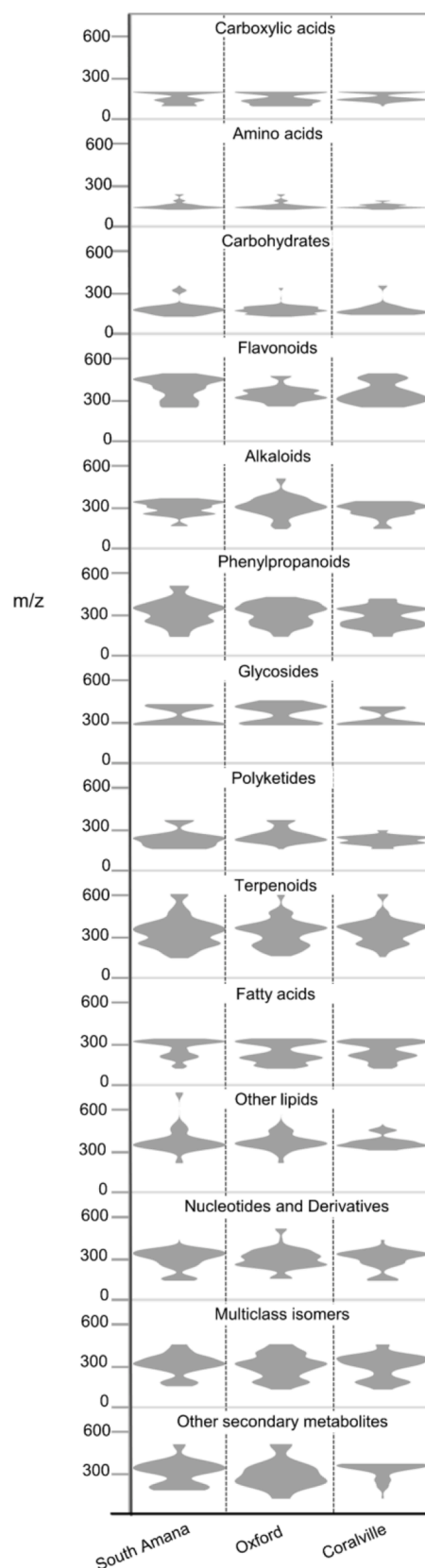


Fig. 4. Spatial dynamics of metabolomics-based profiling of DOM composition along the stream. Violin plots of m/z values of molecular grouping classified by the metabolomics-based approach at the three sampling locations along the stream.

Supplementary Fig. S7). The metabolomics-based analysis determined that this decline was attributed primarily to a decrease in the number of fatty acids (from 99 to 16) at Oxford, while there were minimal changes in the abundance of other lipids (predominantly fatty acyls and sterol lipids) during the storm event (Supplementary Fig. S8).

At Coralville, the abundance of different compounds changed dynamically during the early flush of the storm event (Supplementary Fig. S8). For instance, the number of identified fatty acids fluctuated between 9 and 35 during this period, or between 8 and 32 for nucleotides and associated derivatives (Supplementary Fig. S8). Analysis of within-group dynamics of the metabolites revealed that medium-chain fatty acids (C16-C18; $m/z > 250$) were the most abundant at South Amana, but short-chain fatty acids appeared at Oxford and Coralville (Fig. 4a). Moreover, the abundance of other lipids with m/z of 300–350, corresponding to sterol lipids, remained the same across all three sites. Furthermore, in accordance with increased abundance (from 296 to 372) of phytochemical-like (or keto-aromatic) compounds during the rising limb of the hydrograph followed by a decline during the falling limb (to ~ 330), the metabolomics-based analysis revealed a corresponding pattern for the flavonoids (Supplementary Fig. S7 and S8). Specifically, a shift from high-molecular-weight compounds ($400 < m/z < 500$) to the low-molecular-weight compounds ($250 < m/z < 350$) in the abundance of flavonoids was reminiscent of the breakdown of the glycosyl bond facilitated by glycosyl hydrolases or related enzymes commonly found in streams (Hill et al., 2010; Santos et al., 2019). Such breakdown pattern was also manifested in the disappearance of other glycosides with high m/z ($400 < m/z < 450$) at Coralville (Fig. 4). In addition, a shift in m/z values was also observed for other phenylpropanoids, from $300 < m/z < 400$ to $200 < m/z < 300$ (Fig. 4). Furthermore, a decreased abundance of low-molecular-weight terpenoids ($m/z \sim 250$) downstream also reflected the degradation in the terpenoid group of compounds (Fig. 4). In contrast, the consistent occurrence of similar m/z values for both amino acids and carboxylic acids at all three sites reflected the ubiquity of these compounds in natural waters (Fig. 4) (Mostofa et al., 2013).

To explore further the spatiotemporal molecular variations in the DOM dynamics, we performed principal component analysis (PCA) to evaluate variation in the normalized abundance of chemical formulae in each molecular grouping (using either MSC classification or metabolomics-based analysis) with respect to the different sites, the falling and rising limb of the storm hydrograph, the DOC concentration, and the normalized discharge (Fig. 5). For MSC classification, the first principal component (PC1) explained 55% variations in the abundance of annotated compounds (Fig. 5). This component positively corresponded with phytochemical-like (33% loading), nucleotide-like (28% loading), protein-like (3.2% loading), and carbohydrate-like (0.5% loading) compounds, and negatively corresponded with lipid-like compounds (34% loading) (Fig. 5). The second principal component (PC2) explained 27% of the variations and largely corresponded with protein-like (43% loading), carbohydrate-like (54% loading), and lipid-like (1.6% loading) compounds (Fig. 5). For compounds classified by the metabolomics-based approach, the PC1 and PC2, together, explained 58.6% of the variation (Fig. 5). The PC1 was positively associated with flavonoids (15% loading), other phenylpropanoids (14% loading), and polyketides (comprised of mono and polycyclic aromatics; 7% loading), and negatively corresponded with fatty acids (7.2% loading), other lipids (14% loading) (Fig. 5). The PC2 was positively associated with fatty acids (9% loading) and other lipids (1% loading) (Fig. 5). Notably, for both MSC and metabolomics-based classifications, the PC1 distinguished Coralville (the downstream site) from the other two upstream sites and the PC2 separated South Amana and Oxford (the upstream sites); both the flow stage (i.e., rising and falling limb of hydrograph) and discharge played a small role in explaining the variability of the abundance of molecules while DOC levels correlated negatively with PC1 (Fig. 5).

Therefore, for the compounds classified by either MSC or

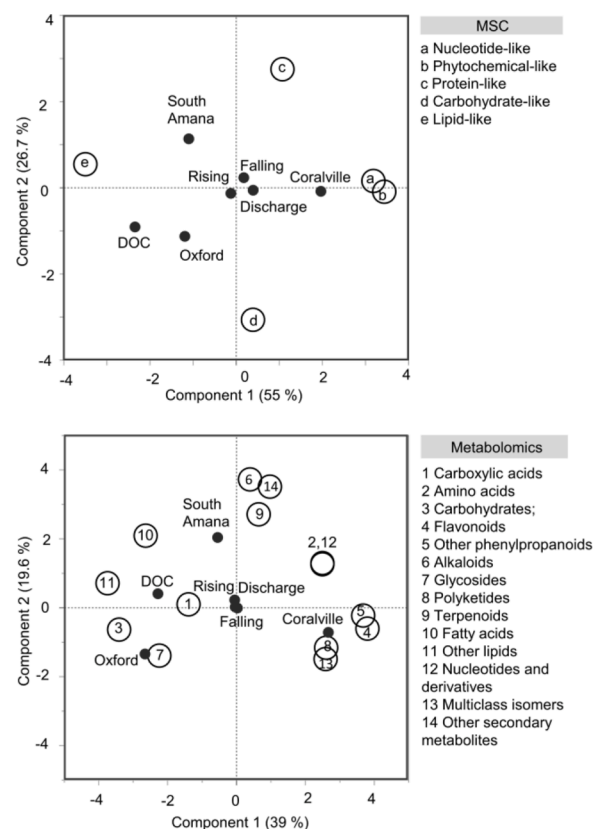


Fig. 5. Correlation of molecules to hydrologic and landscape features of the stream. Biplots of the principal component analysis (performed on% of the total at each site) of the abundance of compounds classified by MSC classification and metabolomics-based classification. Discharge values during the storm were normalized to the maximum discharge observed at each site; “rising” and “falling” indicate the rising and falling limbs of the hydrograph, respectively.

metabolomics-based classification approaches, the PCA revealed a divergence in the molecular composition of the DOC from upstream to downstream sites and highlighted a greater impact of spatial location compared to storm hydrograph in maintaining this divergence (Fig. 5). Firstly, the PCA of compounds categorized by the MSC classification confirmed a transition from a high abundance of lipid-like compounds at South Amana to an enrichment of phytochemical-like (or keto-aromatic) compounds at Coralville (Fig. 5). These changes in the molecular patterns of the DOM could implicate different spatiotemporal contributions of DOM inputs or different transport of the DOM along the stream. Providing high-resolution molecular insights on the identities of the phytochemical-like compounds associated with Coralville, the PCA of the metabolomics-based classification of compounds correlated flavonoids, phenylpropanoids, and both mono and polycyclic aromatics with this downstream location, demonstrating a molecular signature of plant-derived compounds sourced from the different agricultural and forested areas in the watershed (Fig. 5). Secondly, the abundance of N-containing metabolites (amino acids, nucleotides and derivatives) correlated positively with both PC1 ($\sim 6\%$ loading) and PC2 ($\sim 3\%$ loading), consistent with a low percentage of N-containing compounds at Oxford (Fig. 5, Supplementary Fig. S5). Thirdly, glycosides and carbohydrates negatively corresponded with PC1 and PC2 (Fig. 5). Fourthly, the PCA of the metabolites depicted the consistent presence of carboxylic acids at all three sites in the stream (Fig. 5), characterized by a weak correlation with PC1 (2% loading) and no correlation with PC2 (Fig. 5). Taken collectively, the statistical analysis of our data identified the sampling location as the primary determinant of DOM variability in relation to the change in landscape features and in-stream processes at each location.

4. Discussion

Elucidating temporal and spatial changes in the DOM composition of streams and rivers is critical to unraveling the role of this DOM in the cycling and transport fluxes of carbon in a watershed. Towards deciphering the chemical diversity of aquatic DOM underlying these fluxes has been the application of high-resolution mass spectrometry to annotate chemical formulae (D'Andrilli et al., 2013; Gonsior et al., 2011; Lu et al., 2015; Maizel et al., 2017; Sleighter and Hatcher, 2008; Zhang et al., 2018). These chemical formulae are typically analyzed based on stoichiometric ratio relationships to determine their classification with respect to broad molecular categories, thereby resulting in low-resolution elucidation of the molecular patterns in DOM composition. Capturing high-resolution molecular patterns is especially important in streams and rivers wherein DOM composition is highly dynamic due to, beyond the DOM sourcing from the surrounding landscape and in-stream biotic and abiotic processes, hydrologic events that control the load and residence of DOM spatially and temporally (Battin et al., 2009; Bauer and Bianchi, 2012; Creed et al., 2015; Fasching et al., 2020, 2016; Fellman et al., 2009; Giorgio and Pace, 2021; Lu and Liu, 2019; Lynch et al., 2019; Oeurng et al., 2011; Singh et al., 2014; Tanentzap et al., 2019). Here, to advance high-resolution molecular characterization of stream DOM dynamics, we performed a metabolomics-based analysis of DOM from high-frequency sampling at three different sites on a small stream through a storm event. The measured DOC concentrations at Clear Creek were within the DOC range observed in streams and agricultural effluents (Dhillon and Inamdar, 2013; Hassanpour et al., 2017; Wise et al., 2020). Importantly, our metabolomics-based analysis captured different patterns in the molecular composition of the stream DOM at the different sites, which shed light on the spatiotemporal dynamics of DOM cycling and transport beyond what could be achieved solely based on a stoichiometric-based analysis.

Indeed, stoichiometry of chemical formulae can provide valuable preliminary insights into the DOM dynamics in the stream with respect to the extent of oxidation, aromaticity, and heteroatom content (Roth

et al., 2019). Increased heteroatom content of compounds in the downstream location has been previously attributed mainly to autochthonous production (D'Andrilli et al., 2013; Maizel et al., 2017; Rivas-Ubach et al., 2018). Here, our data depicted a time-dependent heteroatom content related to discharge during the storm event, thus illustrating an association between DOM transport and dynamics in heteroatom composition that could be due to transport of diverse DOM derived from both allochthonous and autochthonous production in the stream. In lieu of the widely-used Van Krevelen plots, which may result in inaccurate assignments of biomolecules solely based on the atomic ratios of H/C versus O/C (Lu et al., 2015; Rivas-Ubach et al., 2018), we employed the MSC classification to analyze the abundance of compounds in relation to multiple atomic ratios of O/C, H/C, N/C, and P/C present in the chemical formulae (Rivas-Ubach et al., 2018).

In sum, from upstream to downstream sites, the stoichiometry-based analysis highlighted an enrichment in phytochemical-like compounds, broadly representing keto-aromatic compounds, and depletion of lipid-like and protein-like compounds. The greater abundance of lipid-like and protein-like compounds at the upstream site could be attributed to greater slope upstream to result in greater contribution of groundwater from soil profiles (Frost et al., 2006), enriched in compounds of microbial origins and relatively depleted in plant-derived compounds (Kaiser and Kalbitz, 2012; Roth et al., 2019), due to lower mobility of these compounds adsorbed to soil particles (McDonough et al., 2020). Moreover, the metabolomics-based analysis revealed that the increased abundance of keto-aromatic moieties of relatively lower molecular weights was due to the breakdown of flavonoids and other phenylpropanoids (Fig. 6), which are components of tannins commonly found in stream DOM (Leenheer and Rostad, 2004; Lu and Liu, 2019). Decreasing molecular weight and increasing aromaticity in the outlet of watersheds were previously connected to photo- and microbial degradation (Lambert et al., 2015), and low molecular weight organic compounds were observed previously to tend to be susceptible to degradation (Hodgkins et al., 2016; Roth et al., 2019; Valle et al., 2020).

Furthermore, we determined that spatial changes in the abundance

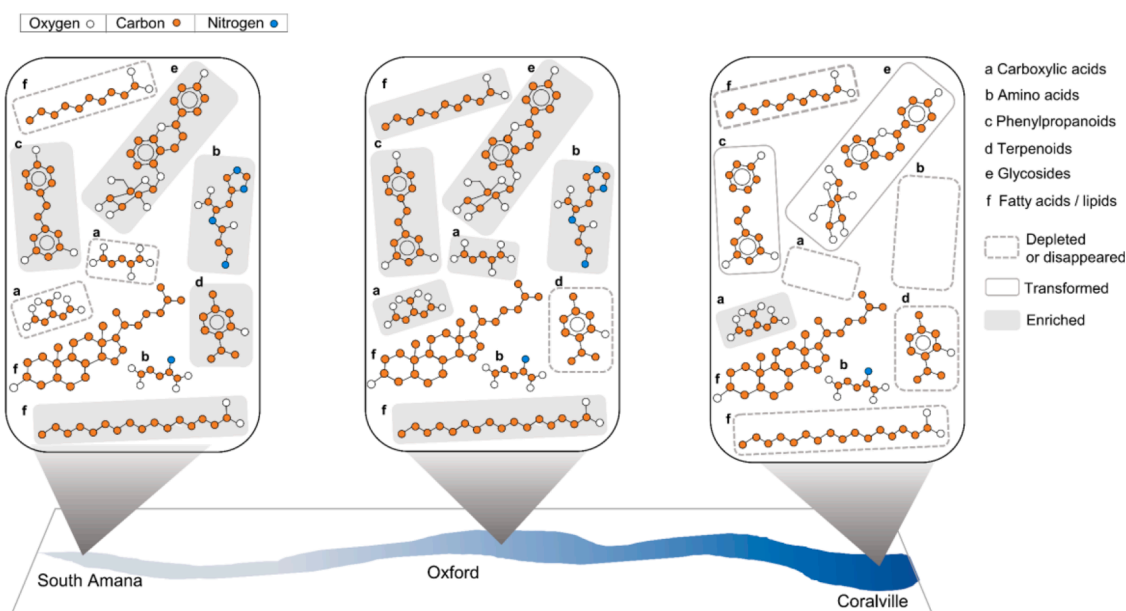


Fig. 6. Conceptual overview of key molecular variability resolved by metabolomics-based analysis of the DOM composition dynamics. Designation of depleted or disappeared (boxed with dashed lines), transformed (boxed with solid line), or enriched (within shaded box) abundance in unique metabolite identities at each site is shown relative to the other two sites. Illustrated are key molecular subgroups that are prevalent in the carbon cycling and transport across the Clear creek corridor going from the upstream site (South Amana) to the midstream site (Oxford), and the downstream site (Coralville): selective enrichment and depletion of carboxylic acids at Oxford and Coralville, disappearance of peptides with persistence of small amino acids at Coralville; breakdown of phenylpropanoids accompanied by increase in relatively smaller aromatic fragments in the downstream sites relative to South Amana, disappearance of terpenoid compounds and breakdown of glycosides at Oxford relative to the upstream sites, and localized enrichment and depletion of short-chain and medium-chain fatty acids but no change in sterol lipids.

of lipid-like compounds were due to localized depletion or enrichment of short-chain and medium-chain fatty acids at each site, whereas sterol lipids remained unchanged throughout (Fig. 6). We also captured persistence of small amino acids in all three sites, disappearance of selected carboxylic acids at the downstream site, and breakdown of glycosides at the downstream sites (Fig. 6). Glycosides are seldom detected in natural water samples, partly due to organic matter extraction methods that hydrolyze the glycosidic bond, but the presence of glycosides in water has ecosystem implications due to their antimicrobial characteristics (Hawkes et al., 2018; Merder et al., 2021). Short-chain carboxylic acids ($C \leq 6$), which are not specifically characterized by stoichiometric ratio relationships despite being a common component of natural organic matter (Adeleke et al., 2017; Aristilde et al., 2017a), are ubiquitous in stream waters as they have high turnover rates (Küsel and Drake, 1998) and can be produced both biotically or abiotically by photo-oxidation (Hassanpour and Aristilde, 2021; Moran and Zepp, 1997).

In headwaters of large watersheds, increased discharge has been reported to promote better connections between river channels, thus leading to a convergent DOM composition with greater similarities due to a greater mixing and availability of nutrients (Lynch et al., 2019). Along a stream, the river continuum concept (Vannote et al., 1980) suggests a continual decrease in the molecular diversity of DOM, although this shift can be influenced by landscape features and anthropogenic activities (Creed et al., 2015; Kamjunke et al., 2019), especially in small streams. Here, in Clear Creek, site-specific inputs from natural areas and tributaries contributed to the increased molecular diversity and DOC concentrations from the upstream site (South Amana) to midstream site (Oxford). However, an inverse association between DOC and the downstream site (Coralville) (Fig. 5), presumably due to both the contribution of a larger area of the watershed and increased retention time at this downstream site, in addition to the immediate urban areas surrounding this site (Frost et al., 2006).

Temporal changes of DOM composition based on our untargeted metabolomics-based analysis at the upstream location of South Amana were similar to those reported for the POC using targeted analysis of biomarker compounds (Blair et al., 2021; Kim et al., 2020). The input sequence of POC into Clear Creek was identified to be fatty acids due to in-channel sources dominated by algae, lignin phenols from surface soils from row crop fields, and finally a complex mixture of upstream sources that became increasingly dominated by channel bank erosional material with transport downstream (Blair et al., 2021; Kim et al., 2020). At the downstream locations (Oxford and Coralville), however, the patterns of DOC and POC differed. Facilitated transport of DOC as a result of the storm flow was not observed at the Oxford site, contrasting the observation that POC at Clear Creek was mobilized during storm events due to increased sediment erosion (Kim et al., 2020). Specifically, in lieu of the reported significant mobilization of POC by erosive forces (Singh Dhillion and Inamdar, 2014), the decrease in compound diversity along with the dilution of the DOC concentration at the Oxford site indicated the lack of sufficient increase in new DOC inputs during the increased runoff, albeit differences between POC and DOC trends may be season-dependent.

We acknowledge that our metabolomics-based approach was focused on small molecules ($m/z < 840$) and comprehensive classification of the annotated chemical formulae is currently not possible due to compound coverage limitation in existing metabolomics databases. Nevertheless, the metabolomics-based analysis presented here spanned a wide variety of compounds belonging to a diverse array of molecular subgroups that are relevant and significant in environmental samples.

5. Conclusion

An important component of carbon cycling in a watershed is the fate of DOM in streams and rivers. The advent of high-resolution mass spectrometry-based techniques has advanced molecular understanding

of the chemical diversity of aquatic DOM. However, stoichiometry-focused analysis of chemical formulae has led to a limited understanding of the dynamic changes in the DOM composition, due to low-resolution molecular classification. Here we employed a metabolomics-based approach to achieve high-resolution fingerprinting of molecular derivatives of plant and microbial metabolism. We obtained spatiotemporal molecular insights into the DOM composition of a stream traversing different landscape types. Interestingly, using either the stoichiometry-based or the metabolomics-based classification, statistical data analysis attributed the dynamic change in the molecular classes or groupings at each site in the stream to carbon flux from the surrounding landscapes combined with in-stream transformations, while the flow stage and the increased discharge played a limited role. However, the metabolomics-based analysis afforded high-resolution evaluation of the spatial divergence in the molecular identities of the compounds, which was not resolved by the stoichiometric analysis. For instance, increased abundance of phytochemical-like compounds was due to the dynamics of specific molecular subgroups captured by metabolomics-based analysis involving increased abundance of polyketides, flavonoids and other phenylpropanoids from upstream to downstream. Through several illustrations of the spatiotemporal molecular variations of the DOM composition in the Clear Creek corridor, we demonstrate here that metabolomics-based analysis can offer additional molecular insights consistent with or unresolved by stoichiometric classification. Thus, the application of such metabolomics-based analysis combined with stoichiometric-based categorization can advance our molecular understanding of the fate of carbon in aquatic environments, a critical component of the global carbon cycle (Lu et al., 2015; Lynch et al., 2019).

Additional information

Supplementary information is available for this article.

CRedit authorship contribution statement

Bahareh Hassanpour: Visualization, Data curation, Formal analysis, Methodology, Conceptualization, Investigation, Software, Writing – original draft, Writing – review & editing. **Neal Blair:** Resources, Writing – review & editing. **Ludmilla Aristilde:** Conceptualization, Formal analysis, Funding acquisition, Resources, Supervision, Visualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Metabolomics LC-HRMS data will be available in MetaboLights repository (www.ebi.ac.uk/metabolights/ MTBLS4164) upon publication under the accession MTBLS4164. Data will also be made available upon request.

Acknowledgements

Funding for this research and postdoctoral funding for B.H. was provided by NSF CAREER Grant (CBET-1653092) awarded to L.A. Support for N.B. was provided by the NSF Grant EAR 2012850, for the project Network Cluster CINet: Critical Interface Network in Intensively Managed Landscapes, a multidisciplinary and multi-institution collaborative effort. Samples were provided by the Intensively Managed Landscape – Critical Zone Observatory project that was supported by the

multi-institutional NSF Grant EAR-1331906. The authors would like to thank Dr. Jean-François Gaillard (Northwestern University) and Ying-qian (Chan) Xiong (Northwestern University) for their assistance with the measurements of DOC concentration.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.watres.2022.118923](https://doi.org/10.1016/j.watres.2022.118923).

References

- Adeleke, R., Nwangburuka, C., Oboirien, B., 2017. Origins, roles and fate of organic acids in soils: a review. *S. Afr. J. Bot.* 108, 393–406. <https://doi.org/10.1016/j.sajb.2016.09.002>.
- Aristilde, L., Guzman, J.F., Klein, A.R., Balkind, R.J., 2017a. Compound-specific short-chain carboxylic acids identified in a peat dissolved organic matter using high-resolution liquid chromatography–mass spectrometry. *Org. Geochem.* 111, 9–12. <https://doi.org/10.1016/j.orggeochem.2017.06.006>.
- Aristilde, L., Reed, M.L., Wilkes, R.A., Youngster, T., Kukurugya, M.A., Katz, V., Sasaki, C.R.S., 2017b. Glyphosate-induced specific and widespread perturbations in the metabolome of soil pseudomonas species. *Front. Environ. Sci.* 5, 34. <https://doi.org/10.3389/fenvs.2017.00034>.
- Battin, T.J., Luyssaert, S., Kaplan, L.A., Aufdenkampe, A.K., Richter, A., Tranvik, L.J., 2009. The boundless carbon cycle. *Nat. Geosci.* 2, 598–600. <https://doi.org/10.1038/ngeo618>.
- Bauer, J.E., Bianchi, T.S., 2012. Dissolved Organic Carbon Cycling and Transformation, Treatise on Estuarine and Coastal Science. Elsevier Inc. <https://doi.org/10.1016/B978-0-12-374711-2.00502-7>.
- Benton, H.P., Want, E.J., Ebbels, T.M.D., 2010. Correction of mass calibration gaps in liquid chromatography-mass spectrometry metabolomics data. *Bioinformatics* 26, 2488–2489. <https://doi.org/10.1093/bioinformatics/btq441>.
- Bhattacharya, R., Osburn, C.L., 2020. Spatial patterns in dissolved organic matter composition controlled by watershed characteristics in a coastal river network: the Neuse River Basin, USA. *Water Res.* 169, 115248. <https://doi.org/10.1016/j.watres.2019.115248>.
- Blair, N.E., Bettis, E.A., Filley, T.R., Moravek, J.A., Papanicolaou, A.N.T., Ward, A.S., Wilson, C.G., Zhou, N., Kazmierczak, B., Kim, J., 2021. The spatiotemporal evolution of storm pulse particulate organic carbon in a low gradient. *Agric. Domin. Watershed Front. Water* 3, 1–17. <https://doi.org/10.3389/frwa.2021.600649>.
- Chu, C., Erickson, P.R., Lundeen, R.A., Stamatelatos, D., Alaimo, P.J., Latch, D.E., McNeill, K., 2016. Photochemical and nonphotochemical transformations of cysteine with dissolved organic matter. *Environ. Sci. Technol.* 50, 6363–6373. <https://doi.org/10.1021/acs.est.6b01291>.
- Creed, I.F., McKnight, D.M., Pellerin, B.A., Green, M.B., Bergamaschi, B.A., Aiken, G.R., Burns, D.A., Findlay, S.E.G., Shanley, J.B., Striegl, R.G., Aulenbach, B.T., Clow, D.W., Laudon, H., McGlynn, B.L., McGuire, K.J., Smith, R.A., Stackpole, S.M., 2015. The river as a chemostat: fresh perspectives on dissolved organic matter flowing down the river continuum. *Can. J. Fish. Aquat. Sci.* 72, 1272–1285. <https://doi.org/10.1139/cjfas-2014-0400>.
- Cyle, K.T., Klein, A.R., Aristilde, L., Martinez, C.E., 2020. Ecophysiological study of paraburkholderia sp. 1N under soil solution conditions: dynamic substrate preferences and characterization of carbon use efficiency. *Appl. Environ. Microbiol.* 1–18. <https://doi.org/10.1128/aem.01851-20>.
- D'Andrilli, J., Foreman, C.M., Marshall, A.G., McKnight, D.M., 2013. Characterization of IHSS Pony Lake fulvic acid dissolved organic matter by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry and fluorescence spectroscopy. *Org. Geochem.* 65, 19–28. <https://doi.org/10.1016/j.orggeochem.2013.09.013>.
- Dhillon, G.S., Inamdar, S., 2013. Extreme storms and changes in particulate and dissolved organic carbon in runoff: entering uncharted waters? *Geophys. Res. Lett.* 40, 1322–1327. <https://doi.org/10.1002/grl.50306>.
- Fasching, C., Akotoye, C., Bižić, M., Fonvielle, J., Ionescu, D., Mathavarajah, S., Zoccarato, L., Walsh, D.A., Grossart, H.P., Xenopoulos, M.A., 2020. Linking stream microbial community functional genes to dissolved organic matter and inorganic nutrients. *Limnol. Oceanogr.* 65, S71–S87. <https://doi.org/10.1002/lno.11356>.
- Fasching, C., Ulseth, A.J., Schelker, J., Steniczka, G., Battin, T.J., 2016. Hydrology controls dissolved organic matter export and composition in an alpine stream and its hyporheic zone. *Limnol. Oceanogr.* 61, 558–571. <https://doi.org/10.1002/lno.10232>.
- Fellman, J.B., Hood, E., Edwards, R.T., D'Amore, D.V., 2009. Changes in the concentration, biodegradability, and fluorescent properties of dissolved organic matter during stormflows in coastal temperate watersheds. *J. Geophys. Res.* 114, G01021. <https://doi.org/10.1029/2008JG000790>.
- Frost, P.C., Larson, J.H., Johnston, C.A., Young, K.C., Maurice, P.A., Lamberti, G.A., Bridgman, S.D., 2006. Landscape predictors of stream dissolved organic matter concentration and physicochemistry in a lake superior river watershed. *Aquat. Sci.* 68, 40–51. <https://doi.org/10.1007/s00027-005-0802-5>.
- Garayburu-caruso, V.A., Danczak, R.E., Stegen, J.C., Renteria, L., McCall, M., Goldman, A.E., Chu, R.K., Toyoda, J., Resch, C.T., Torgeson, J.M., Wells, J., Fansler, S., Kumar, S., Graham, E.B., 2020. Using community science to reveal the global chemogeography of river metabolomes. *Metabolites* 10, 1–20.
- Giorgio, P.A., Pace, M.L., 2021. Relative independence of dissolved organic carbon transport and processing in a large temperate river : the hudson river as both pipe and reactor. *Limnol. Oceanogr.* 53, 185–197.
- Gonsior, M., Peake, B.M., Cooper, W.T., Podgorski, D.C., D'Andrilli, J., Dittmar, T., Cooper, W.J., 2011. Characterization of dissolved organic matter across the subtropical convergence off the South Island, New Zealand. *Mar. Chem.* 123, 99–110. <https://doi.org/10.1016/j.marchem.2010.10.004>.
- Graeber, D., Gelbrecht, J., Pusch, M.T., Anlanger, C., von Schiller, D., 2012. Agriculture has changed the amount and composition of dissolved organic matter in Central European headwater streams. *Sci. Total Environ.* 438, 435–446. <https://doi.org/10.1016/j.scitotenv.2012.08.087>.
- Hassanpour, B., Aristilde, L., 2021. Redox-related metabolic dynamics imprinted on short-chain carboxylic acids in soil water extracts: a 13 C-exometabolomics analysis. *Environ. Sci. Technol. Lett.* 8, 183–191. <https://doi.org/10.1021/acs.estlett.0c00922>.
- Hassanpour, B., Giri, S., Pluer, W.T., Steenhuis, T.S., Geohring, L.D., 2017. Seasonal performance of denitrifying bioreactors in the Northeastern United States: field trials. *J. Environ. Manag.* 202, 242–253. <https://doi.org/10.1016/j.jenvman.2017.06.054>.
- Hawkes, J.A., Patriarca, C., Sjöberg, P.J.R., Tranvik, L.J., Bergquist, J., 2018. Extreme isomeric complexity of dissolved organic matter found across aquatic environments. *Limnol. Oceanogr. Lett.* 3, 21–30. <https://doi.org/10.1002/lo2.10064>.
- Hill, B.H., Elonen, C.M., Jicha, T.M., Bolgrien, D.W., Moffett, M.F., 2010. Sediment microbial enzyme activity as an indicator of nutrient limitation in the great rivers of the upper Mississippi river basin. *Biogeochemistry* 97, 195–209. <https://doi.org/10.1007/s10533-009-9366-0>.
- Hodgkins, S.B., Tfaily, M.M., Podgorski, D.C., McCalley, C.K., Saleska, S.R., Crill, P.M., Rich, V.L., Chanton, J.P., Cooper, W.T., 2016. Elemental composition and optical properties reveal changes in dissolved organic matter along a permafrost thaw chronosequence in a subarctic peatland. *Geochim. Cosmochim. Acta* 187, 123–140. <https://doi.org/10.1016/j.gca.2016.05.015>.
- Hou, T., Berry, T.D., Singh, S., Hughes, M.N., Tong, Y., Thanos Papanicolaou, A.N., Wacha, K.M., Wilson, C.G., Chaubey, I., Filley, T.R., 2018. Control of tillage disturbance on the chemistry and proportion of raindrop-liberated particles from soil aggregates. *Geoderma* 330, 19–29. <https://doi.org/10.1016/j.geoderma.2018.05.013>.
- Kaiser, K., Kalbitz, K., 2012. Cycling downwards - dissolved organic matter in soils. *Soil Biol. Biochem.* 52, 29–32. <https://doi.org/10.1016/j.soilbio.2012.04.002>.
- Kamjunke, N., Hertkorn, N., Harir, M., Schmitt-Kopplin, P., Griebler, C., Brauns, M., von Tümpling, W., Weitere, M., Herzsprung, P., 2019. Molecular change of dissolved organic matter and patterns of bacterial activity in a stream along a land-use gradient. *Water Res.* 164, 114919. <https://doi.org/10.1016/j.watres.2019.114919>.
- Kim, J., Blair, N.E., Ward, A.S., Goff, K., 2020. Storm-induced dynamics of particulate organic carbon in clear creek, iowa: an intensively managed landscape critical zone observatory story. *Front. Water* 2, 1–20. <https://doi.org/10.3389/frwa.2020.578261>.
- Kuhl, C., Tautenhahn, R., Böttcher, C., Larson, T.R., Neumann, S., 2012. CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal. Chem.* 84, 283–289. <https://doi.org/10.1021/ac202450g>.
- Kukurugya, M.A., Mendonca, C.M., Solhtalab, M., Wilkes, R.A., Thannhauser, T.W., Aristilde, L., 2019. Multi-omics analysis unravels a segregated metabolic flux network that tunes Co-utilization of sugar and aromatic carbons in Pseudomonas putida. *J. Biol. Chem.* 294, 8464–8479. <https://doi.org/10.1074/jbc.RA119.007885>.
- Küsel, K., Drake, H.L., 1998. Microbial turnover of low molecular weight organic acids during leaf litter decomposition. *Soil Biol. Biochem.* 31, 107–118. [https://doi.org/10.1016/S0038-0717\(98\)00111-4](https://doi.org/10.1016/S0038-0717(98)00111-4).
- Ladd, M.P., Reeves, D.T., Poudel, S., Iversen, C.M., Wullschlegel, S.D., Hettich, R.L., 2021. Untargeted exometabolomics provides a powerful approach to investigate biogeochemical hotspots with vegetation and polygon type in arctic tundra soils. *Soil Syst.* 5, 10. <https://doi.org/10.3390/soilsystems5010010>.
- Lambert, T., Darchambeau, F., Bouillon, S., Alhou, B., Mbega, J.D., Teodoru, C.R., Nyoni, F.C., Massicotte, P., Borges, A.V., 2015. Landscape control on the spatial and temporal variability of chromophoric dissolved organic matter and dissolved organic carbon in large African rivers. *Ecosystems* 18, 1224–1239. <https://doi.org/10.1007/s10021-015-9894-5>.
- Lu, K., Liu, Z., 2019. Molecular level analysis reveals changes in chemical composition of dissolved organic matter from south texas rivers after high flow events. *front. Mar. Sci.* 6, 1–18. <https://doi.org/10.3389/fmars.2019.00673>.
- Lu, Y., Li, X., Mesfioui, R., Bauer, J.E., Chambers, R.M., Canuel, E.A., Hatcher, P.G., 2015. Use of ESI-FTICR-MS to characterize dissolved organic matter in headwater streams draining forest-dominated and pasture-dominated watersheds. *PLoS ONE* 10, 1–21. <https://doi.org/10.1371/journal.pone.0145639>.
- Lusk, M.G., Toor, G.S., 2016. Dissolved organic nitrogen in urban streams: biodegradability and molecular composition studies. *Water Res.* 96, 225–235. <https://doi.org/10.1016/j.watres.2016.03.060>.
- Lynch, L.M., Sutfin, N.A., Fegh, T.S., Boot, C.M., Covino, T.P., Wallenstein, M.D., 2019. River channel connectivity shifts metabolite composition and dissolved organic matter chemistry. *Nat. Commun.* 10. <https://doi.org/10.1038/s41467-019-08406-8>.
- Maizel, A.C., Li, J., Remucal, C.K., 2017. Relationships between dissolved organic matter composition and photochemistry in lakes of diverse trophic status. *Environ. Sci. Technol.* 51, 9624–9632. <https://doi.org/10.1021/acs.est.7b01270>.
- McDonough, L.K., O'Carroll, D.M., Meredith, K., Andersen, M.S., Brügger, C., Huang, H., Rutledge, H., Behnke, M.I., Spencer, R.G.M., McKenna, A., Marjo, C.E., Oudone, P., Baker, A., 2020. Changes in groundwater dissolved organic matter character in a

- coastal sand aquifer due to rainfall recharge. *Water Res.* 169 <https://doi.org/10.1016/j.watres.2019.115201>.
- Merder, J., Röder, H., Dittmar, T., Feudel, U., Freund, J.A., Gerds, G., Kraberg, A., Niggemann, J., 2021. Dissolved organic compounds with synchronous dynamics share chemical properties and origin. *Limnol. Oceanogr.* 1–16. <https://doi.org/10.1002/lno.11938>.
- Moran, M.A., Zepp, R.G., 1997. Role of photoreactions in the formation of biologically labile compounds from dissolved organic matter. *Limnol. Oceanogr.* 42, 1307–1316. <https://doi.org/10.4319/lno.1997.42.6.1307>.
- Mostofa, K., Yoshioka, T., Abdul Mottaleb, M., Vione, D., 2013. Photobiogeochemistry of Organic Matter, Springer, Environmental Science and Engineering. Springer Berlin Heidelberg, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-642-32223-5>.
- Oeurng, C., Sauvage, S., Coynel, A., Maneux, E., Etcheber, H., Sanchez-Perez, J.-M., 2011. Fluvial transport of suspended sediment and organic carbon during flood events in a large agricultural catchment in southwest France. *Hydrol. Process.* 2378, 2365–2378. <https://doi.org/10.1002/hyp.7999>.
- Pang, Y., Wang, K., Sun, Y., Zhou, Y., Yang, S., Li, Y., He, C., Shi, Q., He, D., 2021. Linking the unique molecular complexity of dissolved organic matter to flood period in the Yangtze River mainstream. *Sci. Total Environ.* 764, 142803 <https://doi.org/10.1016/j.scitotenv.2020.142803>.
- Papanicolaou, A.N.T., Wilson, C.G., Abaci, A., Elhakeem, M., Skopec, M., 2010. SOM loss and soil quality in the clear creek, IA. *J. Iowa Acad. Sci. JIAS* 116, 14–26. Vol.
- Phungsai, P., Kurisu, F., Kasuga, I., Furumai, H., 2016. Molecular characterization of low molecular weight dissolved organic matter in water reclamation processes using Orbitrap mass spectrometry. *Water Res.* 100, 526–536. <https://doi.org/10.1016/j.watres.2016.05.047>.
- Remucal, C.K., Cory, R.M., Sander, M., McNeill, K., 2012. Low molecular weight components in an aquatic humic substance as characterized by membrane dialysis and Orbitrap mass spectrometry. *Environ. Sci. Technol.* 46, 9350–9359. <https://doi.org/10.1021/es302468q>.
- Riedel, T., Zark, M., Vähätalo, A.V., Niggemann, J., Spencer, R.G.M., Hernes, P.J., Dittmar, T., 2016. Molecular signatures of biogeochemical transformations in dissolved organic matter from ten world rivers. *Front. Earth Sci.* 4, 1–16. <https://doi.org/10.3389/feart.2016.00085>.
- Rivas-Ubach, A., Liu, Y., Bianchi, T.S., Tolić, N., Jansson, C., Paša-Tolić, L., 2018. Moving beyond the van Krevelen diagram: a new stoichiometric approach for compound classification in organisms. *Anal. Chem.* 90, 6152–6160. <https://doi.org/10.1021/acs.analchem.8b00529>.
- Roth, V., Lange, M., Simon, C., Hertkorn, N., Bucher, S., Goodall, T., Griffiths, R.I., Mellado-vázquez, P.G., Mommer, L., Oram, N.J., Weigelt, A., Dittmar, T., Gleixner, G., 2019. Persistence of dissolved organic matter explained by molecular changes during its passage through soil. *Nat. Geosci.* 12, 755–761. <https://doi.org/10.1038/s41561-019-0417-4>.
- Santos, C.D., Sarmiento, H., de Miranda, F.P., Henrique-Silva, F., Logares, R., 2019. Uncovering the gene machinery of the Amazon River microbiome to degrade rainforest organic matter. *Microbiome* 30 (8), 151. <https://doi.org/10.1186/s40168-020-00930-w>.
- Schilling, K.E., Helmers, M.J., 2010. Advanced bash-scripting guide an in-depth exploration of the art of shell scripting table of contents. *Hydrol. Process.* 2274, 2267–2274. <https://doi.org/10.1002/hyp>.
- Schilling, K.E., Jindal, P., Basu, N.B., Helmers, M.J., 2012. Impact of artificial subsurface drainage on groundwater travel times and baseflow discharge in an agricultural watershed, Iowa (USA). *Hydrol. Process.* 26, 3092–3100. <https://doi.org/10.1002/hyp.8337>.
- Singh Dhillon, G., Inamdar, S., 2014. Storm event patterns of particulate organic carbon (POC) for large storms and differences with dissolved organic carbon (DOC). *Biogeochemistry* 118, 61–81. <https://doi.org/10.1007/s>.
- Singh, S., Inamdar, S., Mitchell, M., McHale, P., 2014. Seasonal pattern of dissolved organic matter (DOM) in watershed sources: influence of hydrologic flow paths and autumn leaf fall. *Biogeochemistry* 118, 321–337. <https://doi.org/10.1007/s10533-013-9934-1>.
- Sleighter, R.L., Hatcher, P.G., 2008. Molecular characterization of dissolved organic matter (DOM) along a river to ocean transect of the lower Chesapeake bay by ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Mar. Chem.* 110, 140–152. <https://doi.org/10.1016/j.marchem.2008.04.008>.
- Smith, C.A., O'Maille, G., Want, E.J., Qin, C., Trauger, S.A., Brandon, T.R., Custodio, D. E., Abagyan, R., Siuzdak, G., 2005. METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* 27, 747–751. <https://doi.org/10.1097/01.ftd.0000179845.53213.39>.
- Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., Siuzdak, G., 2006. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* 78, 779–787. <https://doi.org/10.1021/ac051437y>.
- Stutter, M.I., Richards, S., Dawson, J.J.C., 2013. Biodegradability of natural dissolved organic matter collected from a UK moorland stream. *Water Res.* 47, 1169–1180. <https://doi.org/10.1016/j.watres.2012.11.035>.
- Tanentzap, A.J., Fitch, A., Orland, C., Emilson, E.J.S., Yakimovich, K.M., Osterholz, H., Dittmar, T., 2019. Chemical and microbial diversity covary in fresh water to influence ecosystem functioning. *Proc. Natl. Acad. Sci. USA* 116, 24689–24695. <https://doi.org/10.1073/pnas.1904896116>.
- Tautenhahn, R., Bottcher, C., Neumann, S., 2008. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinform.* 9, 1–16. <https://doi.org/10.1186/1471-2105-9-504>.
- Valle, J., Harir, M., Gonsior, M., Enrich-Prast, A., Schmitt-Kopplin, P., Bastviken, D., Hertkorn, N., 2020. Molecular differences between water column and sediment pore water SPE-DOM in ten Swedish boreal lakes. *Water Res.* 170, 115320 <https://doi.org/10.1016/j.watres.2019.115320>.
- Vannote, R.L., Minshall, G.W., Cummins, K.W., Sedell, J.R., Cushing, C.E., 1980. The river continuum concept. *Can. J. Fish. Aquat. Sci.* 37, 130–137. <https://doi.org/10.1139/f80-017>.
- Leenheer, J.A., Rostad, C.E., 2004. Tannins and terpenoids as major precursors of Suwannee River fulvic acid, Scientific Investigations Report. US. Geological Survey, Reston, Virginia. 10.3133/sir20045276.
- W.N. Venables, D.M. Smith, and the R Core Team, 2020. An introduction to R; notes on R: a programming environment for data analysis and graphics version 3.6.3 (2020-02-29) 3.
- Vione, D., Colombo, N., Said-Pullicino, D., Bocchiola, D., Confortola, G., Salerno, F., Viviano, G., Fratianni, S., Martin, M., Godone, D., Freppaz, M., 2021. Seasonal variations in the optical characteristics of dissolved organic matter in glacial pond water. *Sci. Total Environ.* 759, 143464 <https://doi.org/10.1016/j.scitotenv.2020.143464>.
- Wise, J.L., Horn, D.J.V., Diefendorf, A.F., Regier, P.J., Lowell, T.V., Dahm, N., 2020. Dissolved organic matter dynamics in storm water runoff in a dryland urban region. *J. Arid Environ.* 165, 55–63. <https://doi.org/10.1016/j.jaridenv.2019.03.003>.
- Yang, L., Chang, S.W., Shin, H.S., Hur, J., 2015. Tracking the evolution of stream DOM source during storm events using end member mixing analysis based on DOM quality. *J. Hydrol.* 523, 333–341. <https://doi.org/10.1016/j.jhydrol.2015.01.074>.
- Zhang, L., Wang, S., Yang, J., Xu, K., 2018. In-depth molecular characterization and biodegradability of water-extractable organic nitrogen in Erhai Lake sediment. *Environ. Sci. Pollut. Res.* 25, 19779–19789. <https://doi.org/10.1007/s11356-018-2122-z>.