






ARTICLE

<https://doi.org/10.1038/s41467-019-08406-8>

OPEN

# River channel connectivity shifts metabolite composition and dissolved organic matter chemistry

Laurel M. Lynch <sup>1,2</sup>, Nicholas A. Sutfin <sup>3</sup>, Timothy S. Fegell<sup>4</sup>, Claudia M. Boot <sup>2,5</sup>, Timothy P. Covino <sup>2,6</sup> & Matthew D. Wallenstein <sup>2,6</sup>

Biogeochemical processing of dissolved organic matter (DOM) in headwater rivers regulates aquatic food web dynamics, water quality, and carbon storage. Although headwater rivers are critical sources of energy to downstream ecosystems, underlying mechanisms structuring DOM composition and reactivity are not well quantified. By pairing mass spectrometry and fluorescence spectroscopy, here we show that hydrology and river geomorphology interactively shape molecular patterns in DOM composition. River segments with a single channel flowing across the valley bottom export DOM with a similar chemical profile through time. In contrast, segments with multiple channels of flow store large volumes of water during peak flows, which they release downstream throughout the summer. As flows subside, losses of lateral floodplain connectivity significantly increase the heterogeneity of DOM exported downstream. By linking geomorphologic landscape-scale processes with microbial metabolism, we show DOM heterogeneity increases as a function of fluvial complexity, with implications for ecosystem function and watershed management.

<sup>1</sup>Section of Soil and Crop Sciences, College of Agriculture and Life Sciences, Cornell University, Ithaca, NY, USA. <sup>2</sup>Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, CO, USA. <sup>3</sup>Department of Earth, Environmental, and Planetary Sciences, Case Western Reserve University, Cleveland, OH, USA. <sup>4</sup>United States Forest Service, Rocky Mountain Research Station, Fort Collins, CO, USA. <sup>5</sup>Central Instrument Facility, Department of Chemistry, Colorado State University, Fort Collins, CO, USA. <sup>6</sup>Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO, USA. Correspondence and requests for materials should be addressed to L.M.L. (email: [lml262@cornell.edu](mailto:lml262@cornell.edu))

**D**issolved organic matter (DOM) accounts for the vast majority of reactive carbon flowing through aquatic ecosystems<sup>1,2</sup>. Given the central role of DOM in global carbon cycling, nutrient export, and food web dynamics<sup>3,4</sup>, unraveling processes controlling the flux and transformation of DOM in watersheds is essential to understanding potential changes resulting from land use and climate change. Headwater rivers exert disproportionate control on fluvial biogeochemistry<sup>5</sup> by setting initial conditions that cascade to the entire watershed<sup>4</sup> and by providing a critical source of chemical energy to downstream ecosystems. However, current state of the sciences lacks a quantitative understanding of the relationship between seasonal hydrology and geomorphic complexity in structuring DOM composition and reactivity.

In mountainous watersheds, the majority of DOM enters the fluvial network in small headwater tributaries<sup>6</sup>. By some estimates, half of this carbon is mineralized through microbial and photochemical degradation<sup>7</sup>, while the fraction remaining within the watershed is retained on floodplains, exported further downstream, or generated in situ through autochthonous production or heterotrophic fractionation<sup>2,8,9</sup>. River geomorphic complexity—the physical heterogeneity in river channel geometry and planform<sup>10</sup>—is a critical factor controlling the residence time and concentration of DOM residing within fluvial networks<sup>11,12</sup>. This is particularly evident in mountainous channel networks, where rivers alternate between segments with multiple channels of flow (complex or multi-thread) and segments with a single channel of flow (simple) across the valley bottom<sup>13</sup>. Here we show that channel complexity is a significant driver of DOM composition, which varies as a function of lateral river-floodplain hydrologic connectivity.

Stable multi-thread channels develop in broad, relatively unconfined valleys of Rocky Mountain National Park, where channel-spanning logjams and beaver (*Castor canadensis*) activity dissipate transport energy and create sites of flow diversion<sup>13,14</sup>. These multi-thread channels serve as retention zones that enhance the transient storage of DOM<sup>15</sup> and may increase geo-physical opportunities for microbial metabolism<sup>3</sup>. Flume studies have shown that microbial diversity and uptake of chemically complex DOM is positively correlated with physical complexity<sup>16,17</sup>. Diversification of metabolic pathways in complex channels may thus translate into more extensive DOM processing prior to downstream export. In contrast, single-thread channel segments commonly occur in steep-gradient valleys and are characterized by high transport capacities and limited sediment, water, and fluvial carbon storage<sup>9</sup>.

Many once complex, multi-thread segments of wide valleys have been transformed to simple, single-thread segments resulting from flow regulation, land-use changes, and beaver removal<sup>9</sup>. Ensuing loss of geomorphic complexity results in rapid export of DOM through the landscape, potentially minimizing microbial transformation of DOM and limiting river-floodplain hydrologic connectivity, and floodplain carbon storage<sup>9</sup>. While geomorphic complexity appears to be an important driver of DOM composition, limited analytical resolution has precluded detailed investigation<sup>18</sup>. As a result, it remains unclear whether decomposition results in the convergence of DOM molecular composition toward a core set of metabolites, or persistence of the original, chemically diverse molecular profile<sup>19</sup>. Resolution of this debate is essential in determining how DOM is cycled within a landscape.

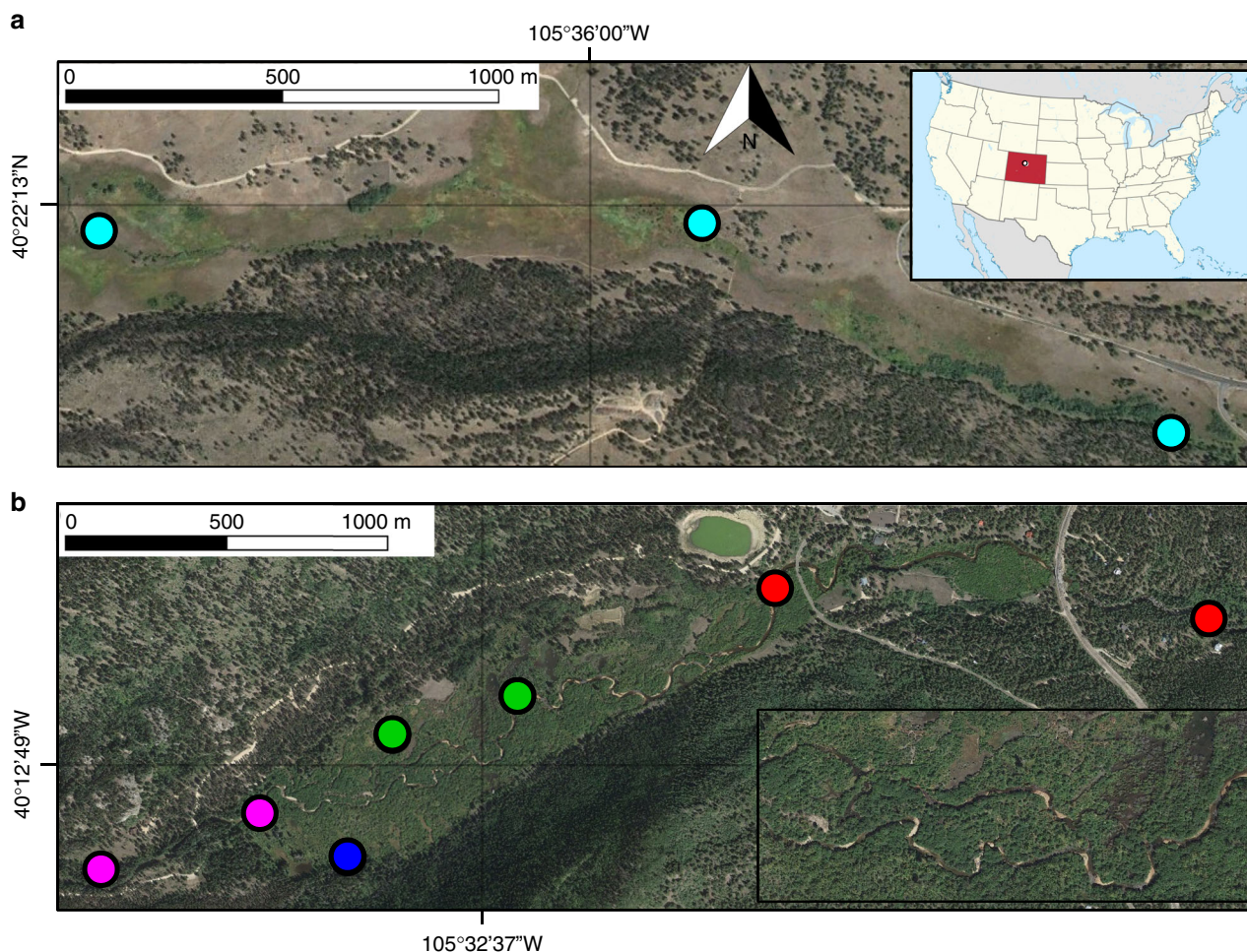
Strong fluctuations in river flows, resulting from seasonal snowmelt or stormflow events, may interact with geomorphic complexity to control molecular patterns of DOM composition. Elucidating the influence of environmental drivers on DOM composition remains challenging because of tremendous chemical heterogeneity<sup>18</sup> and the complexity of underlying spatio-temporal mechanisms influencing its composition. Mass spectrometry has vastly improved our understanding of the composition and cycling of DOM compounds varying in physicochemical reactivity. While fluorescence, nuclear magnetic resonance, pyrolysis gas chromatography mass spectrometry (GC-MS), and Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) approaches are routinely used to probe DOM composition, they group DOM complexity into broad functional groups or compound classes. In contrast, electron ionization GC-MS (EI GC-MS) can identify individual metabolites from complex mixtures of environmental DOM<sup>20</sup>. This detailed characterization of DOM allows inferences about the underlying patterns of microbial metabolism that structure the reactivity and fate of DOM flowing throughout the fluvial network.

In this study, we collected surface and hyporheic water samples to monitor changes in fluvial chemistry through space (spanning 3 km of valley length in two watersheds with river networks of differing geomorphic complexity) and time (eight time points from high to low river flows spanning May to October) (Table 1 and Supplementary Figure 1). We selected two river segments (simple network and complex network), with nearly identical climate regimes and underlying geology that differ substantially in channel complexity. Three sites within the simple network were sampled. Seven sites within the complex network were sampled, including: a beaver pond that has limited surface water

**Table 1 Physical characteristics of sampling sites**

	Simple	Upper	Meadow complex	Lower	Pond
Watershed	UBM	NSV	NSV	NSV	NSV
Channel planform	Single-thread	Single-thread	Multi-thread	Single-thread	Pond
Lateral river-floodplain connectivity	Low	Low	High	Low	None
Longitudinal connectivity	Continuous	Continuous	Intermittent	Continuous	None
Number of sites	3	2	2	2	1
Watershed area (m)	21.6	88.9	88.9	88.9	88.9
Channel width (m)	0.71	12.2	14.5	20	12.8
Valley width (m)	163.9	59.6	246.8	27.5	246.9
Confinement	0.004	0.204	0.059	0.725	0.052
Elevation (m)	2,572.51	2,561.08	2,544.33	2,519.76	2,545.85

These are located in UBM (simple network) and NSV (complex network) rivers, Rocky Mountain National Park, Colorado. Channels with a confinement value < 0.2 are considered unconfined while those with a confinement value ≥ 0.2 are considered confined. Confined is defined as the ratio of the channel width versus the valley width  
UBM Upper Beaver Meadow, NSV North Saint Vrain



**Fig. 1** Site map of the simple and complex river channel networks sampled for this study. **a** From the simple network (Upper Beaver Meadows; UBM), we present data from three single-channel sites (light blue symbols). **b** From the complex network (North Saint Vrain Creek; NSV), we present data from two main channel sites in the upper confined segment (pink symbols), two side-channel features (green symbols), and a disconnected beaver pond (blue symbol) in the active beaver meadow complex, and two main channel sites in the lower confined segment (red symbols). Symbol colors match Fig. 3. The headwaters of UBM and NSV are located in Rocky Mountain National Park, Colorado, USA. Map data ©2018 Google

connectivity to the main channel, and confined upstream and downstream (simple morphology) river segments that bracket the main stem and side-channel features within a wide-valley, multi-thread beaver meadow complex (Fig. 1). We used complementary approaches, including EI GC-MS and fluorescence spectroscopy (excitation-emission matrices), to quantify spatial and temporal differences in DOM molecular composition. We also determined the potential role of geomorphic complexity, hydrologic connectivity, and river discharge in structuring DOM molecular composition. We posit that river segments with high geomorphic complexity increase the chemical diversity of DOM exported downstream, and maximal chemical diversity occurs during high-flow periods and declines over time as flows subside.

## Results

**Influence of channel connectivity on DOM composition.** The variability in DOM chemical composition within the complex river network increased as connectivity with the main channel declined through the summer. In contrast, the chemical composition of DOM exported through the simple network was not influenced by seasonal changes in river flows. Therefore, for each sampling point we used an average of the three sampling locations along the simple network for subsequent statistical analyses.

Bulk carbon concentrations collected from all sampling locations were significantly different across time and space ( $p < 0.01$ ,  $F_{8,39} = 3.55$ ; Table 2). Of all the potential explanatory geomorphic and environmental variables tested, only valley confinement was a significant predictor of DOC concentrations ( $p < 0.001$ , Supplementary Table 1), which were higher during intermediate flows than peak or low flow conditions. Within the complex network, DOC concentrations were significantly higher in the beaver pond than all other sites ( $p < 0.01$ ) but were not different between upper and lower confined segments (Table 2). We also observed an effect of landscape position on hyporheic carbon and nitrogen concentrations, such that sites with lower transport energy (beaver pond and beaver meadow) had higher carbon and nitrogen concentrations than freely flowing reaches ( $F_{8,39} = 2.44$ ;  $p < 0.05$ ).

Total dissolved nitrogen (TDN) concentrations were twice as high in the beaver pond relative to all other sites ( $p < 0.01$ ); hyporheic nitrogen concentrations decreased across all sites throughout the season ( $p < 0.01$ ). Regressing TDN against discharge revealed nitrogen concentrations were highest during peak river flows ( $R^2 = 0.46$ ). Total C:N varied spatially and temporally ( $F_{8,39} = 4.22$ ; Table 2); relative nitrogen availability was higher in the beaver pond and simple channel than the beaver meadow complex and downstream confined channel

**Table 2 Biogeochemical characteristics of surface water and hyporheic sediments**

ID	DOC	TDN	NO <sub>3</sub> <sup>-</sup>	NH <sub>4</sub> <sup>+</sup>	Hyporheic DOC	Hyporheic TDN	Shannon-Weiner
Landscape position							
Simple	4.00 (0.53)	0.25 (0.03)	0.08 (0.02)	0.07 (0.01)	4.81 (0.54)	0.35 (0.05)	3.45 (0.09)
Pond	9.90 (1.95)	0.61 (0.15)	0.03 (0.01)		8.10 (1.49)	0.68 (0.23)	
Upper	2.70 (0.55)	0.33 (0.05)	0.11 (0.01)		3.41 (0.65)	0.58 (0.17)	
Meadow	4.27 (0.49)	0.39 (0.05)	0.08 (0.02)		8.64 (2.29)	1.20 (0.29)	
Lower	3.42 (0.71)	0.31 (0.04)	0.09 (0.02)		4.41 (1.02)	0.73 (0.21)	
Sampling period							
1: May 1	6.49 (0.66)	0.49 (0.07)	0.11 (0.01)	0.17 (0.06)	6.20 (1.76)	2.13 (0.20)	3.45 (0.12)
2: May 14	5.17 (0.67)	0.35 (0.02)	0.08 (0.02)	0.08 (0.01)		1.28 (0.47)	
3: Jun 8	4.72 (0.29)	0.47 (0.04)	0.11 (0.03)	0.08 (0.01)		0.54 (0.06)	
4: Jun 30	3.58 (0.49)	0.38 (0.04)	0.06 (0.02)	0.08 (0.01)		0.49 (0.08)	
5: Jul 21	4.32 (1.75)	0.32 (0.06)	0.04 (0.01)	0.06 (0.01)		0.54 (0.11)	
6: Aug 17	8.22 (3.28)	0.49 (0.15)	0.03 (0.01)	0.05 (0.00)		0.37 (0.08)	
7: Sep 21	4.57 (2.68)	0.23 (0.11)	0.04 (0.01)	0.05 (0.00)		0.42 (0.09)	
8: Oct 28	4.64 (2.49)	0.43 (0.28)	0.06 (0.01)	0.05 (0.00)		0.31 (0.10)	
Source of variance							
LP	**	*	***	ns	*	**	ns
SP	**	ns	*	***	ns	***	ns

These are organized by landscape position and sampling period (discharge) at Upper Beaver Meadows (simple) and North Saint Vrain (pond, upper, meadow, and lower). The average followed by the standard error ( $\pm$ 1 S.E.) in parentheses for dissolved C, N (DOC, TDN), inorganic N (NO<sub>3</sub><sup>-</sup>, NH<sub>4</sub><sup>+</sup>), and a metabolite chemodiversity index (Shannon-Weiner). Aside from the unitless chemodiversity index, all values are reported in mg/L. The level of significance from the two-way analysis of variance model for samples arranged by LP and SP are reported as \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , or nonsignificant (ns). For ns comparisons, a master mean, integrated across landscape position or sampling period, is provided. Specific comparisons were explored using Tukey HSD t-tests with Satterthwaite-Welch approximations for degrees of freedom (reported in the Results section)

DOC dissolved organic carbon, TDN total dissolved nitrogen, LP landscape position, SP sampling period

( $p < 0.05$ ). The fluorescence characteristics of ultraviolet-visible (UV-VIS) DOM (FDOM) exhibited strong temporal dynamics but were not influenced by geomorphic complexity (Supplementary Table 2). We observed the lowest contribution of simple aromatic proteins (regions I and II) and soluble microbial-type proteins (region IV, also interpreted as microbial exudates) during peak river flows, with the contribution of these regions to total FDOM doubling throughout the summer (region I:  $F_{8,39} = 2.84$ ; Fig. 2a). In contrast, the contribution of fulvic (region III)- and humic (region V)-type acids to FDOM decreased in relative intensity by 60% throughout the season, except for a highly enriched signal exported during peak discharge (region III:  $F_{8,39} = 2.84$ ; Fig. 2b). Regressing region I against region III revealed a strong, negative correlation ( $R^2 = 0.89$ ), suggesting high levels of autochthonous productivity occur during periods of low allochthonous subsidy (Supplementary Figure 2). Discharge alone was a significant predictor of regions I ( $p < 0.01$ ,  $R^2 = 0.27$ ) and III ( $p < 0.05$ ,  $R^2 = 0.22$ ), supporting the control of seasonality on FDOM composition (Supplementary Table 1).

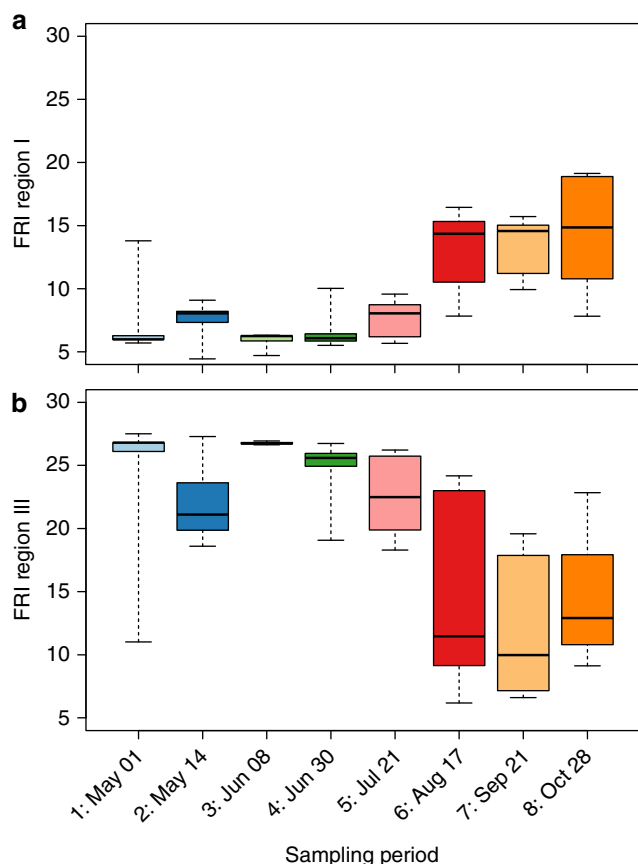
**Metabolite distribution.** We observed a total of 2472 mass spectral features using untargeted EI GC-MS. We resolved these features using cluster analysis, resulting in 259 unique compounds that each consisted of between 3 and 57 mass spectral fragments (Supplementary Figure 3)<sup>21</sup>. We were able to annotate and classify 10% of these compounds<sup>22–24</sup>, which primarily consisted of sugars, organic acids, lipids, and lignin-derived aromatics (Table 3). Due to limitations in current mass spectral libraries we were not able to annotate the majority of compounds driving sample separation across geomorphic complexity and sampling period. However, we were able to use changes in metabolite composition to separate samples in ordination space.

Over half of all profiled metabolites (both annotated and not) exhibited significant separation across regions of varying geomorphic complexity (Supplementary Figure 4a). Combined, partial least squares discriminant analysis (PLS-DA) components 1 and 2 explained 43% of the variance, and retention of four axes maximized classification performance ( $R^2 = 0.69$ ; coefficient of

variation (CV) accuracy = 0.56). As PLS-DA classification can be susceptible to overfitting, we used  $Q^2$  scores to estimate the predictive ability of the model through cross-validation<sup>25</sup>. While our strongly positive score ( $Q^2 = 0.43$ ) suggests the model is not overfitted<sup>26</sup> we also include results from two unsupervised approaches (principal components analysis (PCA), Supplementary Figure 5, and random forest, Supplementary Figure 6) that display similar trends. The five primary VIP compounds, or metabolites that contributed most to sample separation across geomorphic complexity, had slightly lower parent ion  $m/z$  values (352.0) than average (402.2), but low annotation scores preclude meaningful metabolite assignments (Supplementary Table 3).

Spatial variability in metabolite composition among upstream and downstream reaches was distinctly different between the simple and complex meadow sites. The beaver pond and simple system, which represent end-members of hydrologic (dis) connectivity, had divergent metabolite profiles (Fig. 3a), indicating a strong influence of geomorphic complexity and hydrologic connectivity on metabolite distribution. Within the complex network, the upper and lower confined reaches showed minimal overlap in metabolite distribution, while the beaver meadow complex, situated in physical space between the two sites, shared metabolites with each. The upper confined sites and beaver meadow complex sites were more dissimilar than the lower confined site, which contained metabolites very similar to, and within the range of, the meadow site (Fig. 3d). We observed a similar pattern when considering only annotated metabolites. The five primary VIP compounds contributing to metabolite separation across geomorphic complexity were related to byproducts of decomposition, including nonanoic and dodecanoic acids (medium-chain saturated fatty acids), glyceric acid (a three-carbon sugar acid), and isovanillic acid (a derivative of lignin degradation) (Table 3)<sup>23</sup>. These compounds had the highest scores in the beaver pond and simple network, the lowest scores in the upper confined channel, and relatively low scores within the beaver meadow complex and lower confined channel (Supplementary Figure 7a).

Metabolites also separated across sampling period (Fig. 3b), although only 40% exhibited significant temporal variability



**Fig. 2** Boxplots represent seasonal distributions of excitation-emission matrix spectra (EEMS) regions. EEMS region I: simple aromatic proteins (**a**), and EEMS region III: fulvic-type acids (**b**) across sampling period ( $n = 49$ ) within the complex river network. Regions II and IV (simple aromatic proteins and soluble microbial byproducts) follow similar seasonal patterns as region I, while region III and V (humic-type acids) are related. Each box spans the interquartile range; whiskers extend to the minimum and maximum of the distribution

(Supplementary Figure 4b). Together, the first two PLS-DA components explained 45% of the variance, with maximal classification performance achieved with five axes ( $R^2 = 0.96$ ; CV accuracy = 0.44,  $Q^2 = 0.80$ ). The five primary annotated VIP compounds were typically enriched early in the season and included lumichrome (a flavin pigment), glycerol (a three-carbon sugar acid), 2-hydroxypyridine (potentially involved in DNA synthesis), and several polyethylene glycols (ether compounds) (Table 3 and Supplementary Figure 7)<sup>23</sup>.

Samples collected during the rising limb and peak discharge (Fig. 3c) were composed of significantly different metabolites than samples collected during baseflow (Fig. 3d), with typically lower parent ion  $m/z$  values. Early in the season, samples did not separate by landscape position ( $p = 0.81$ ), suggesting high hydrologic connectivity during high-flow periods homogenizes the DOM pool and exports a core set of metabolites (Fig. 3c). In contrast, we observed significant sample separation across geomorphic complexity during late season sampling ( $p = 0.01$ ), suggesting metabolite profiles diverge during low flow periods (Fig. 3d).

**Environmental drivers of metabolite diversity.** We identified six environmental drivers associated with metabolite composition (Fig. 4) using constrained analysis of principal coordinates<sup>27</sup>. Within ordination space, principal coordinate 1 explained 43% of

variation across the dataset and principal coordinate 2 explained an additional 30%. Dissolved organic carbon (DOC) concentrations emerged as the primary factor influencing metabolite composition and was negatively correlated with discharge (Fig. 4), where high discharge exported metabolites with relatively lower parent ion  $m/z$  values than low flow periods. Additionally, metabolites clustering with higher DOC concentrations had higher parent ion  $m/z$  values than those associated with proteins (excitation-emission matrix spectra (EEMS) region I,  $p < 0.003$ ) and snowpack extent ( $p < 0.006$ ), indicating lower chemical reactivity. Samples clustering with proteins were typically collected later in the summer, when low flows reduced hydrologic connectivity across the landscape. These patterns suggest more complex river corridors export more chemically diverse, reactive metabolites from multi-thread channels relative to their single-thread counterparts.

## Discussion

Our results suggest that fluvial chemistry and the molecular composition of DOM exported through mountainous watersheds are strongly influenced by seasonality and channel geomorphology. Broad indices of DOM composition, including optical fluorescence (UV-VIS) and DOC and TDN concentrations, were strongly related to in-stream flows. While these bulk chemical approaches revealed the influence of seasonality, we found the application of non-targeted metabolomics—a higher-resolution approach capable of profiling both known and unknown metabolites<sup>21,22</sup>—helped resolve the influence of landscape complexity on DOM chemistry. We suggest linking complex environmental drivers to ecosystem metabolism is critical in understanding network-level energetics, an area of significant research interest<sup>3,11,16,18,28</sup>.

We observed the highest and most variable contributions of EEMS regions III and V to the total FDOM pool prior to peak discharge, suggesting the landscape is hydrologically connected during snowmelt. These fluorescent regions are associated with aromatic, terrestrial-derived organics<sup>29,30</sup> and linked to DOM mobilization across the terrestrial-aquatic interface<sup>5,31</sup>. In sub-alpine watersheds, initial changes in river water chemistry are driven by snowmelt activation of shallow soil flowpaths<sup>32</sup>. Spring thaw rapidly increases landscape hydrologic connectivity, releasing metabolites that have accumulated in isolated soil pores throughout the winter<sup>33</sup>. Materials transported from near-surface soil compartments are enriched in reactive plant-derived compounds, especially phenolics and other lignin derivatives<sup>34</sup>. We observed high concentrations of metabolites with relatively lower parent ion  $m/z$  values, which support activation of shallow flowpaths during snowmelt and export of plant-derived materials<sup>35</sup>, including syringol and isovannillinic acid, which are associated with lignin-type derivatives<sup>23</sup>. Snowmelt water that intersects deeper soil flowpaths mobilizes compounds enriched in microbial-derivatives and nitrogen-bearing compounds<sup>36</sup>. Thus, river water chemistries exported in the spring may reflect activation of diverse soil flowpaths<sup>31</sup>, however, rapid export during peak flows may result in mixing and convergence of the chemical profile. Although high concentrations of potentially labile DOM are exported with snowmelt, cold temperatures, rapid flow velocities, and relatively low solar insolation likely limit in-stream metabolism<sup>5</sup>.

During peak flows, we observed strongly homogenized fluvial and metabolite chemistries, and hyporheic scouring in single-channel segments. We propose these simplified reaches typify the “pulse-shunt” concept proposed by Raymond et al.<sup>37</sup>, where high concentrations of terrestrial-derived DOM mobilized during snowmelt can bypass local metabolism due to rapid increases in

**Table 3 Annotated metabolites collected from subalpine watersheds**

Compound	Score	Ion intensity	<i>m/z</i>	Retention time (s)	Molecular weight (g/mol)	KEGG classification
Lyxose	2	1.07E + 08	73.09	905.41	150.13	Carbohydrate (pentose sugar)
Dodecaethylene glycol	2	1.04E + 08	117.11	1,048.61	546.65	Ether (polyethylene glycols)
2-hydroxy-pyridine	2	9.15E + 07	117.12	883.29	95.10	Pyridinones
Glycerol	2	2.89E + 08	131.12	842.33	92.09	Sugar alcohol
Triethylene glycol	2	5.48E + 08	144.14	157.86	150.17	Ether (polyethylene glycols)
Galactose	2	7.25E + 07	144.14	259.73	180.16	Hexose (carbohydrate)
Tripropylene glycol mono-n-butyl ether	2	2.02E + 07	151.08	360.78	248.36	Ether (polyethylene glycols)
Syringic acid	2	1.93E + 08	152.07	191.13	198.17	Gallic acid and derivatives
Lumichrome	2	2.37E + 08	187.16	755.85	242.23	Flavin pigment
Palmitic acid	2	5.26E + 06	189.09	344.35	256.42	Long-chain fatty acid
Arabinofuranose	2	3.57E + 07	217.12	457.45	150.13	Pentose (carbohydrate)
Isovanillic acid	2	3.20E + 06	219.14	848.21	168.15	P-methoxybenzoic acids and derivatives
Heptaethylene glycol	2	1.33E + 06	295.16	738.69	326.39	Ether (polyethylene glycols)
Propylene glycol	2	1.59E + 07	297.10	551.73	76.09	Secondary alcohol
Benzoic acid	2	5.88E + 06	317.16	563.13	122.12	Benzoic acid
Coumaric acid	2	3.02E + 06	361.17	1,033.60	164.05	Hydroxycinnamic acids and derivatives
Sucrose	2	2.79E + 07	371.29	848.38	342.30	O-glycosyl compounds (disaccharides)
Dodecanoic acid	2	1.36E + 06	455.29	1,008.71	200.32	Medium-chain fatty acid (lipids)
Stearic acid	2	1.00E + 06	473.23	994.22	284.48	Long-chain fatty acid (lipids)
Glyceric acid	2	1.15E + 06	502.20	1,018.32	106.08	Sugar acids and derivatives
Nonanoic acid	2	9.43E + 05	506.07	983.12	158.23	Medium-chain fatty acid (lipids)
Thiazole	2	1.07E + 06	521.79	943.04	85.13	Thiazole
Decaethylene glycol	2	6.69E + 05	589.26	1,024.46	458.55	Ether (polyethylene glycols)

The ion intensity and *m/z* values are reported for the largest ion fragment within each clustered spectrum (where each spectrum represents a single TMS-derived metabolite). Lower parent ion *m/z* values indicate higher compound volatility and are typically associated with faster elution. Spectra were clustered using RamClust. Annotation scores were acquired by querying NIST Standard Reference and KEGG databases, and classified using ClassyFire, a taxonomic database. InChI identifiers are reported in Supplementary Table 4  
TMS trimethylsilyl, KEGG Kyoto Encyclopedia of Genes and Genomes

river velocity. Systems with more homogenous flows have also been linked with lower microbial biodiversity and DOM uptake relative to sites with longer residence times that facilitate opportunities for microbial metabolism<sup>17</sup>. The lack of upstream DOM processing in more simple, homogenous valley corridors has implications for headwater carbon storage and water quality as well as the redistribution of carbon and nutrient cycling to downstream, higher-order rivers.

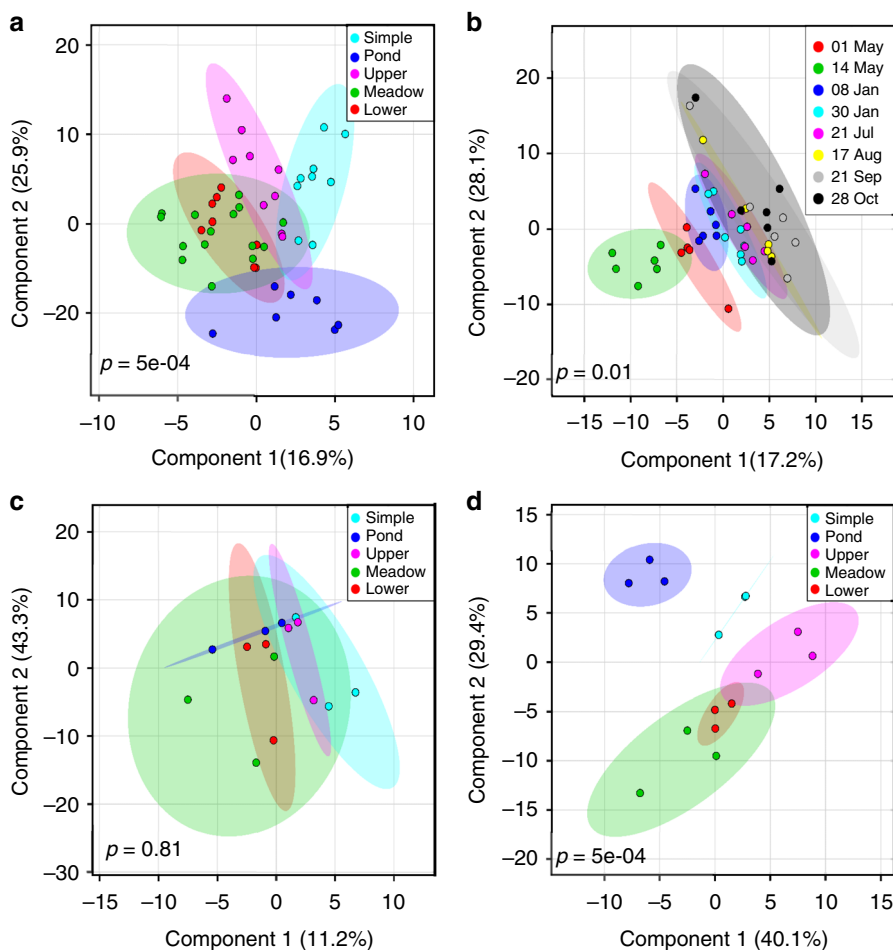
Like many mountainous headwaters, the complex North Saint Vrain (NSV) network exhibits pronounced downstream variations in valley width and river planform<sup>6</sup> that moderate pulse-shunt dynamics even during high-flow periods. Specifically, single-thread segments alternate with broad valley segments containing multi-thread channel morphologies that can promote hydrological buffering by dissipating transport energy<sup>9</sup>. Wegener et al.<sup>11</sup> showed that the multi-thread beaver meadow complex exhibits variable sink-source dynamics as a function of flow. During high-flow periods, water and entrained nutrients are distributed laterally in floodplain surface water bodies and are gradually released during low-flow periods. The release of these components subsidizes downstream metabolism, integrating ecosystem metabolism with landscape connectivity<sup>38</sup>.

Throughout the summer (June–August), we observed significant reductions in stream velocity and increasing indicators of autochthonous productivity, suggesting greater opportunities for in-stream metabolism and decreased lateral connectivity across the terrestrial-aquatic interface<sup>5</sup>. Increasing relative contributions of soluble microbial-type proteins and decomposition byproducts (fluorescence regional integration (FRI) regions I, II, and IV) to total FDOM were particularly pronounced in the beaver pond and the beaver meadow complex, which are less shaded by

riparian vegetation and mountains than more confined sites. In addition, these low-flow sites have increased DOM residence times, and are typically warmer<sup>11</sup>, enhancing geophysical opportunities for metabolism relative to high-flow sites<sup>3</sup>. Within the beaver pond, we observed high concentrations of saturated medium-chain (9:0 and 10:0) and long-chain (16:0 and 18:0) fatty acids, both of which have been linked to fungal and bacterial detritivores<sup>39</sup>, and can serve as significant energy sources for aquatic heterotrophs.

During late summer baseflow (September–October) we observed significant landscape fragmentation as side and main channels within the beaver meadow complex became hydrologically disconnected<sup>9</sup>. Resulting fluctuations in redox gradients have been linked to the activation of metabolic pathways utilizing alternate terminal electron acceptors to degrade DOM<sup>2,40,41</sup>. Together with greater light exposure and photo-oxidative transformation<sup>42,43</sup>, these conditions could contribute to observed increases in chemical diversity with hydrograph recession. Alternatively, shifts in DOM composition could be explained by changes in hydrologic flow path, which transition from soil matrix to groundwater inputs during baseflow<sup>5,44</sup>. However, groundwater sources are typically associated with lower molecular weight, less chemically diverse DOM<sup>31,34</sup>. Thus, while the increased contribution of soluble proteins (FRI regions I, II, and IV) to total FDOM could be explained by decreased loading of matric porewater from upland sources (enriched in FRI Regions III and V), the overall increase in metabolite diversity suggests microbial metabolism and photo-oxidation partially regulates increases in DOM heterogeneity as flows subside.

While seasonally driven increases in photo-oxidation<sup>2</sup> could explain some of the divergence in DOM composition with



**Fig. 3** Partial least squares discriminant analysis (PLS-DA) plot of metabolites. Metabolites are clustered by landscape position (**a**), sampling period (**b**), rising limb, including only May 01, May 15, and June 08 (**c**), or baseflow, including only Aug 17, Sep 21, and Oct 28 (**d**). Each circle represents a sample and shaded ellipses represent 95% confidence intervals for each a priori cluster. Ellipse overlap signifies no significant difference between clusters. For each PLS-DA model, 2000 permutations were generated to test whether the classification systems (landscape position and sampling period) were significantly better ( $p < 0.05$ ) than those arising through random chance

hydrograph recession, results presented here appear to support a shift from anabolic assimilation toward catabolic metabolism<sup>19</sup>. Specifically, we did not observe an increase in common photo-products, such as acetic, citric, pyruvic, or formic acids, which may have been rapidly metabolized<sup>45,46</sup>. Instead, declines in nutrient concentrations and hydrologic connectivity—assessed as a reduction in discharge—were associated with warmer temperatures in disconnected slough features and higher ecosystem respiration within the beaver meadow complex<sup>11</sup>. These patterns suggest nutrient-limited microbial communities cycle DOM less efficiently during warmer, low-flow periods, promoting carbon release to the atmosphere rather than incorporation in microbial biomass<sup>47</sup>. Unlike anabolism, which promotes chemical convergence through assimilation, catabolic respiration preserves the inherent complexity of DOM<sup>19</sup>. Thus, we revise our initial hypothesis and propose instead that carbon chemistry in subalpine rivers tends toward chemostasis during high-flow periods, when landscape connectivity and hydrologic drivers shunt DOM through the fluvial network, and hydrologic fragmentation shifts the balance from anabolic to catabolic metabolism, thereby increasing carbon complexity (Fig. 5).

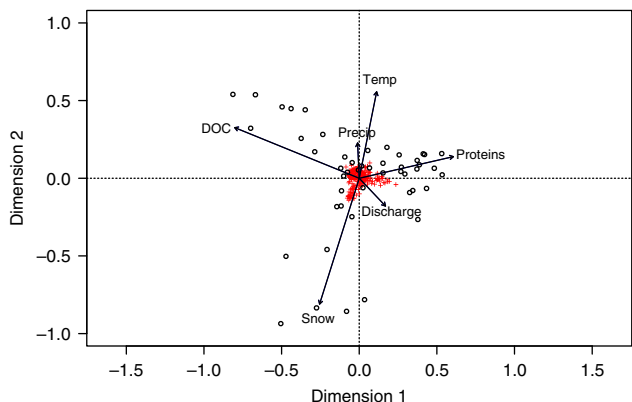
Although still in its nascence, the ability to infer microbial functionality from metabolomics is a promising avenue of research<sup>18,23</sup>. EI GC-MS provides fragmentation spectra useful

for identifying individual metabolites, but databases are sparse for environmental applications<sup>48</sup>. For example, products of secondary metabolism (aromatics, alkaloids, terpenoids, glycosides, phenolics, and lignins) are not represented as well as those of primary metabolism (amino acids, sugars, organic acids, and peptides)<sup>20</sup>. We found the majority of compounds separating samples by landscape position had relatively high parent ion values  $m/z$  ( $>300$ ), potentially indicative of secondary metabolites and other compounds with relatively high molecular weights.

The role of secondary metabolites in structuring network-level energetics is an area of considerable interest<sup>35,49–51</sup>. The environmental ubiquity of these structurally complex compounds suggests they confer a competitive advantage upon the organisms producing them. Otherwise, the pressures of Darwinian natural selection would preclude their synthesis<sup>49</sup>. Mounting evidence suggests secondary metabolites are variously involved in gene expression and cellular growth, with the potential to create more beneficial environmental conditions (for instance, by complexing iron and nutrients)<sup>50</sup>. Others have suggested the production of diverse metabolites facilitates microbial adaptation to complex environmental stressors<sup>51</sup>, which are pronounced in subalpine watersheds, particularly under a changing climate. Although we could not annotate many of these compounds, by pairing metabolomics with fluorescence spectroscopy, we were able to link

microbial metabolism and landscape-scale processes to make inference about how underlying metabolic pathways structure local patterns of DOM chemistry throughout the fluvial network.

The ability of subalpine watersheds to store and process DOM is threatened by widespread and systematic channel simplification, resulting from land use changes, flow regulation, instream beaver and wood removal, and river channelization<sup>11,52,53</sup>. In addition to loss of geomorphic complexity, changes in hydrologic

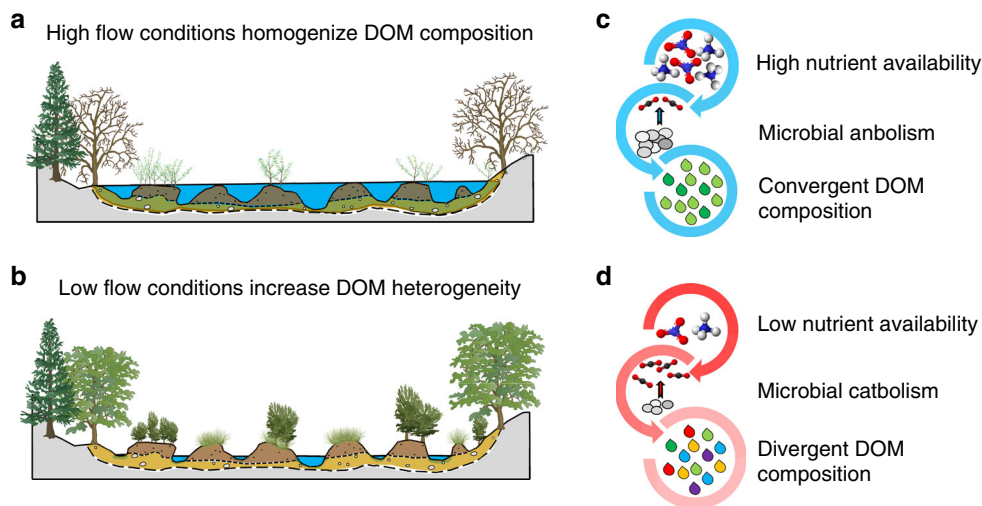


**Fig. 4** Multivariate analysis of metabolites and environmental drivers using capscale. Constrained analysis of principal coordinates (capscale) ordinations are based on Bray-Curtis dissimilarity, computed from normalized peak intensities, and used to test for an association of metabolite composition (as derived from electron ionization gas chromatography mass spectrometry) with spatial, climate, and fluvial characteristics. Red crosses represent individual metabolites, black circles represent samples, black vectors represent environmental variables with a significance level of  $<0.05$  ( $p$ -values based on analysis of variance with 1000 permutations; the relationship between all potential predictor variables and metabolites is shown in Supplementary Figure 8). DOC, dissolved organic carbon; Precip, month-averaged precipitation; Temp, month-averaged temperature; Proteins, EEMS region I; Discharge, daily average; Snow, daily snow depth at Wild Basin

regime associated with reduced snowpack and stochasticity in the timing and magnitude of runoff have the potential to greatly alter the spatio-temporal dynamics of hydrologic connectivity at the terrestrial-aquatic interface. These watersheds are also experiencing dramatic shifts in biogeochemical function, including increased nitrogen deposition from intensified industrial and agricultural production<sup>54</sup>, episodic acidification<sup>55</sup>, and reductions in DOM and water subsidies from shrinking glaciers<sup>28</sup>. The combination of these factors reduces ecosystem function, with implications for human health<sup>56</sup>, the value of water and storage-associated flow regimes<sup>57</sup>, fate and transport of point and non-point source pollution<sup>53</sup>, and storm water management<sup>58</sup>. Here we suggest subalpine systems are particularly sensitive to channel simplification because the loss of geomorphic complexity reduces metabolic opportunities for DOM processing, autochthonous productivity, hydrologic connectivity, and the maintenance of flow path and DOM diversity. We also suggest seasonal reductions in hydrologic connectivity, exacerbated by reductions in snowpack<sup>59</sup>, will heighten catabolic metabolism by reducing nutrient availability in disconnected features. The propagation of these novel chemistries fuels downstream metabolism, tying local biogeochemical processing to network-level energetics.

## Methods

**Site description.** NSV and Beaver Brook are headwater rivers located in Rocky Mountain National Park, Colorado, USA. Both rivers have nearly identical climate regimes and underlying geology, consisting of a granitic core<sup>60</sup>, but differ substantially in channel planform complexity. The approximately 4-km study reach along NSV exhibits pronounced downstream variations in valley geometry and channel complexity (Table 1). Channel planform along NSV alternates between single and multi-thread channel segments as a result of beaver activity<sup>6</sup>. The total drainage area of the NSV study reach is  $\sim 90$  km<sup>2</sup> with an elevation of  $\sim 2560$  m. Vegetation is typical of the Colorado Front Range, where aspen (*Populus tremuloides*) and willow stands (*Salix spp.*) dominate the riparian zone, and Engelmann spruce (*Picea engelmannii*), lodge pole pine (*Pinus contorta*), and subalpine fir (*Abies lasiocarpa*) colonize upland sites<sup>11</sup>. Approximately 3-km long, the Beaver Brook study site is herein referred to as Upper Beaver Meadow (UBM). This site was once a complex channel segment but has been abandoned by beavers since the early twentieth century, resulting in channel simplification<sup>13</sup>. As a result, a narrow (1–2 m wide), single-thread channel has incised (1–2 m) into legacy beaver meadow sediment<sup>61</sup>. Within the portion of Rocky Mountain National Park where these sites are located, mean annual temperature is 5 °C with a summer average of 14 °C,



**Fig. 5** Conceptual model relating hydrologic connectivity and metabolite composition. During peak discharge (**a**), high flows limit geophysical opportunities for microbial metabolism. Landscape fragmentation during summer baseflow (**b**) increases water residence time and microbial processing of dissolved organic matter (DOM) and available nutrients. With sufficient nutrient availability, microbial anabolism dominates (**c**), causing DOM chemistry to converge toward a core set of metabolites. As nutrient availability decreases in isolated side-channel and slough features, catabolic pathways activate (**d**) and metabolite diversity increases. Seasonal changes in flow path, greater autochthonous productivity, and increased exposure to sunlight may also contribute to divergent DOM composition under low flow conditions. Ball and stick models depict carbon dioxide (CO<sub>2</sub>), nitrate (NO<sub>3</sub><sup>-</sup>), and ammonium (NH<sub>4</sub><sup>+</sup>) molecules, where C atoms are shown in charcoal, O atoms in red, N atoms in blue, and H atoms in light gray



and mean annual precipitation is 861 mm, with 451 mm contributed as snowfall (Copeland Lake SNOTEL site # 412, 2621 m elevation, 40° 7' 22.0794"N by 105° 20' 26.5194"W).

Study reaches along UBM and NSV consist of three longitudinally adjacent river segments (Fig. 1; Table 1). At NSV we sampled a multi-thread, active beaver meadow complex bounded by relatively confined upstream and downstream single-thread segments. An upstream and downstream transect bounded each of the three NSV subreaches, and the NSV beaver meadow included an additional transect at the middle of the subreach. The NSV beaver meadow complex has the highest planform complexity, with multiple water impoundments behind beaver dams, as well as side channels and slough features that progressively disconnect from the main channel with hydrograph recession<sup>11</sup>. Samples were also collected from a beaver pond located along the upstream transect within the NSV beaver meadow, which has no surface water connectivity and sits ~3 m higher in elevation than the main channel. UBM is a relatively unconfined single channel segment with three sample locations separated by ~1500 m, and there are no multi-thread channels within the study reach.

The selection of these sites allowed us to assess the temporally dynamic role of lateral river-floodplain connectivity (high, low, and none) and longitudinal hydrologic connectivity (continuous, intermittent, and none) on fluvial biogeochemistry (Table 1). At each sampling site, we recorded GPS coordinates and physical characteristics including river (or pond) width and depth. River discharge was recorded hourly at UBM from June 11 through September 30, 2015 and at NSV from April 1 through August 30, 2015 (reported by Wegener et al.<sup>11</sup>).

**Sample collection and processing.** We collected paired surface water and hyporheic sediment samples eight times between May and September 2015 ( $n = 48$ , Supplementary Figure 1). Surface water samples were collected into sterile 50 mL amber borosilicate bottles and hyporheic sediments were collected into sterile 50 mL centrifuge vials. Additional large-volume surface water samples were collected into 10 L sterile containers and prepared for EI GC-MS analysis (see below). All samples were filtered within six hours of collection through 0.7  $\mu\text{m}$  glass fiber filters pre-combusted at 400 °C (Whatman GF/F). Samples analyzed for total DOC and TDN were acidified to pH 3 and refrigerated until analysis on a Shimadzu TOC-L (Shimadzu Corporation Columbia, MD). Subsamples analyzed for inorganic nitrogen ( $\text{NO}_3^-$  and  $\text{NH}_4^+$ ) were frozen until analysis on an Alpkem flow solution IV automated wet chemistry system (O.I. Analytical College Station, TX).

**Excitation-emission matrices.** We used excitation-emission matrices to study the seasonal variability of UV-VIS FDOM characteristics<sup>62</sup>. To reduce inner-filter effects and normalize organic carbon concentrations, we diluted filtered samples to 5 mg C/L before analyzing them on an Aqualog spectrofluorometer with a xenon excitation source (Horiba-Jobin Yvone Scientific Edison, NJ). We set excitation and emission slits to a 3 nm band-pass, and incrementally increased wavelength in 3 nm steps from 200 to 800 nm. A sealed cuvette was used as a blank and analyzed between every twenty samples to correct for instrument drift. Following spectral analysis, we corrected each sample for inner-filter effects, masked Rayleigh scatter using first and second grating orders, and normalized each sample spectra by the blank<sup>63</sup>. We quantified percent intensities for regions of the EEMS using the FRI approach and separated five spectral regions as outlined by Chen et al.<sup>29</sup> using Matlab version R2016b. EEMS regions I and II are related to simple aromatic proteins (tyrosine and tryptophan-like), region III to fulvic acid-type materials, region IV to soluble microbial byproduct-type materials, and region V to humic acid-type organics<sup>29</sup>. The FRI approach is better suited to capture the underlying heterogeneity of aromatic DOM as it quantifies entire regions of wavelength-dependent fluorescent intensities rather than utilizing only several data points per spectrum<sup>29</sup>.

**Electron ionization gas chromatography-mass spectrometry.** We used EI GC-MS to generate mass spectral fragments, permitting partial structural identification of metabolites present in each sample<sup>20</sup>. We concentrated the total organic carbon in each pre-filtered 10 L surface water sample using four Bond Elut PPL cartridges pre-conditioned with high-performance liquid chromatography (HPLC)-grade methanol (Agilent Technologies). Following organic carbon concentration, each cartridge was extracted with 10 mL of HPLC-grade methanol into sterile borosilicate vials and evaporated under pure  $\text{N}_2$  gas. Samples were prepared for GC-MS analysis at the Proteomics and Metabolomics Facility, Colorado State University. Samples were re-suspended in 5 mL of methanol and vortexed for 30 s. Next, 1 mL aliquots were collected and centrifuged for 10 min at 15,000  $\times g$  and 4 °C, and then 50  $\mu\text{L}$  of *N*-methyl-*N*-trimethylsilyltrifluoroacetamide and 1% trimethylchlorosilane were added to volatilize nonpolar metabolites. Sample supernatants were incubated for 30 min at 60 °C and then centrifuged for five minutes at 3000  $\times g$ . Next, 80  $\mu\text{L}$  aliquots were transferred to a glass GC-MS autosampler vial and injected into a Trace GC Ultra coupled to a Thermo ISQ mass spectrometer (Thermo Scientific) in a 1:10 split ratio. A 30 m TG-5MS column (Thermo Scientific, 0.25 mm i.d., 0.25  $\mu\text{m}$  film thickness) with a 1.2 mL/min helium gas flow rate separated metabolites. Masses between 50–650 mass to charge ( $m/z$ ) were scanned at 5 scans/s after electron impact ionization. Quality control samples were injected after every six samples. A matrix of molecular features, defined by

retention time and  $m/z$ , were generated using XCMS software in R. We grouped all spectral features based on an in-house clustering tool, RAMClustR, which uses spectra based coelution and covariance across the full dataset to group features<sup>21</sup>. Peak areas for each spectral feature were condensed into a single value by adjusting the weighted mean of all spectral features for each compound. Compounds were annotated based on spectral matching to NISTv12, Golm, Metlin, Massbank, and in-house metabolite databases (Table 3)<sup>21</sup>. Two common contaminants, gallic acid and silanol, were also removed prior to subsequent analyses. We procured the InChI code for each annotated metabolite from PubChem and used ClassyFire to identify its chemical taxonomy<sup>23</sup> (Supplementary Table 4). Attempts were made to identify unknown metabolites contributing to sample separation in PLS-DA ordination space using Competitive Fragmentation Modeling for Metabolite Identification<sup>24</sup>. However, fragmentation patterns between the unknown and candidate metabolites were not robust; we include metabolite information and Jaccard scores in Supplementary Table 3 (confidence of assignment < 20% for all unknown compounds) and representative spectral similarity files in Supplementary Figure 9.

**Statistical analysis.** We used one-way analysis of variance (ANOVA) models to identify main effects of landscape position and collection period (Julian date) on each dependent variable of interest using the *lmtest* package<sup>64</sup> in R version 3.3.1<sup>65</sup>. We explored mean differences within each dependent variable using two-sample *t*-tests with unequal variances and Satterthwaite's approximation for degrees of freedom using the *lsmeans* package in R<sup>66</sup>. Data were log-transformed when necessary to meet assumptions of normality, evaluated with Shapiro-Wilk tests. Model residuals were tested for constant and homogenous variance using Q-Q and residual versus fitted plots. Outliers were identified using Cook's distance and removed if  $D_i \geq 1$ .

We used multiple linear regression analysis to identify relationships between our dependent variable of interest (DOC, EEMS region I, or EEMS region III) and potential predictor variables. Due to high correlations among many possible explanatory variables analyses (with  $r \geq 0.6$ ), only those most consistent with past work or theory are included in our regression analyses (with  $r \geq 0.6$ ). A correlation matrix of all possible predictors is included in Supplementary Figure 10. Potential variables for the initial regression analysis included channel width, river discharge, precipitation, temperature, valley confinement, and valley topographic aspect. Stepwise multiple linear regressions were conducted following Sutfin and Wohl<sup>18</sup>, using the *step* function in the R *stats* package<sup>65</sup>. We normalized DOC using the boxcox power transformation in the R *MASS* package, with  $\lambda = -0.343434$ . Boxcox transformations yielded normally distributed errors, assessed with Shapiro-Wilk tests of the residuals (0.85), and inspection of Q-Q plots and histograms of model residuals.

We used MetaboAnalyst 3.0 for detailed metabolite processing and analysis<sup>25</sup>. We removed non-informative metabolites using interquartile range filtering, and normalized the data using a log transformation and Pareto variance scaling, which are common data pretreatment methods<sup>67</sup>. We identified significantly different metabolites using one-way ANOVA and Tukey's HSD post hoc analyses and determined whether samples clustered by categorical factors (landscape position or sampling period), using PLS-DA, a method commonly applied to metabolomic and chemometric datasets<sup>68</sup>. We performed PLS-DA regressions in MetaboAnalyst 3.0, which uses the *pls* function in the R *pls* package<sup>69</sup>. We assessed the performance of each generated model using classification and 10-fold cross-validation and used prediction accuracy during training and B/W permutation tests to explore relationships between our metabolite distribution (predictor variables) and hypothetical classification system (landscape position or sampling period). For each PLS-DA model, we generated 2000 permutations to test whether these classification systems were significantly better than those arising through random chance<sup>68</sup>. We also used two unsupervised methods, PCA and random forest<sup>70</sup>, to model the predictive capacity of our two classification systems.

We used constrained analysis of principal components (capscale), an ordination method in the R *vegan* package that permits non-Euclidean dissimilarity indices<sup>27</sup>, to test for an association of metabolite composition (as derived from EI GC-MS) with spatial, climate, and fluvial characteristics. We included 25 potential environmental drivers in our ordination; a full capscale model displaying the association of all potential predictor variables with metabolite composition is provided in Supplementary Figure 8. We computed Bray-Curtis dissimilarities from normalized metabolite peak intensities and fit our environmental drivers to each ordination, with *p*-values calculated over 999 permutations. We extracted sample and metabolite information from the first two canonical axes to test for an association between metabolite characteristics ( $m/z$ , retention time, abundance), and retained ( $p < 0.01$ ) environmental drivers. We used an independent non-metric multidimensional scaling ordination to validate retention of the selected environmental drivers (stress: 0.05).

## Data availability

All data used in this analysis are available through FigShare, an open-source data repository (<https://doi.org/10.6084/m9.figshare.7015190>).

Received: 9 May 2018 Accepted: 8 January 2019

Published online: 28 January 2019

## References

- Battin, T. J. et al. The boundless carbon cycle. *Nat. Geosci.* **2**, 598–600 (2009).
- Evans, C. D. et al. Variability in organic carbon reactivity across lake residence time and trophic gradients. *Nat. Geosci.* **10**, 832–835 (2017).
- Battin, T. J. et al. Biophysical controls on organic carbon fluxes in fluvial networks. *Nat. Geosci.* **2**, 595–595 (2009).
- Cole, J. J. et al. Plumbing the global carbon cycle: integrating inland waters into the terrestrial carbon budget. *Ecosystems* **10**, 171–184 (2007).
- Fasching, C., Ulseth, A. J., Schelker, J., Steniczka, G. & Battin, T. J. Hydrology controls dissolved organic matter export and composition in an Alpine stream and its hyporheic zone. *Limnol. Oceanogr.* **61**, 558–571 (2016).
- Wohl, E., Dwire, K., Sutfin, N., Polvi, L. & Bazan, R. Mechanisms of carbon storage in mountainous headwater rivers. *Nat. Commun.* **3**, 1263 (2012).
- Aufdenkampe, A. K. et al. Riverine coupling of biogeochemical cycles between land, oceans, and atmosphere. *Front. Ecol. Environ.* **9**, 53–60 (2011).
- Sutfin, N. A. & Wohl, E. Substantial soil organic carbon retention along floodplains of mountain streams. *J. Geophys. Res. Earth Surf.* **122**, 1325–1338 (2017).
- Sutfin, N. A., Wohl, E. E. & Dwire, K. A. Banking carbon: a review of organic carbon storage and physical factors influencing retention in floodplains and riparian ecosystems. *Earth Surf. Process. Landf.* **41**, 38–60 (2016).
- Livers, B. & Wohl, E. Sources and interpretation of channel complexity in forested subalpine streams of the Southern Rocky Mountains. *Water Resour. Res.* **52**, 3910–3929 (2016).
- Wegener, P., Covino, T. & Wohl, E. Beaver-mediated lateral hydrologic connectivity, fluvial carbon and nutrient flux, and aquatic ecosystem metabolism. *Water Resour. Res.* **53**, 4606–4623 (2017).
- Dick, J. J., Tetzlaff, D., Birkel, C. & Soulsby, C. Modelling landscape controls on dissolved organic carbon sources and fluxes to streams. *Biogeochemistry* **122**, 361–374 (2015).
- Wohl, E. Landscape-scale carbon storage associated with beaver dams. *Geophys. Res. Lett.* **40**, 3631–3636 (2013).
- Polvi, L. E. & Wohl, E. Biotic drivers of stream planform. *Bioscience* **63**, 439–452 (2013).
- Gooseff, M. N., Hall, R. O. & Tank, J. L. Relating transient storage to channel complexity in streams of varying land use in Jackson Hole, Wyoming. *Water Resour. Res.* **43**, 1–10 (2007).
- Besemer, K., Singer, G., Hödl, I. & Battin, T. J. Bacterial community composition of stream biofilms in spatially variable-flow environments. *Appl. Environ. Microbiol.* **75**, 7189–7195 (2009).
- Singer, G., Besemer, K., Schmitt-Kopplin, P., Hödl, I. & Battin, T. J. Physical heterogeneity increases biofilm resource use and its molecular diversity in stream mesocosms. *PLoS One* **5**, 1–11 (2010).
- Kellerman, A. M., Dittmar, T., Kothawala, D. N. & Tranvik, L. J. Chemodiversity of dissolved organic matter in lakes driven by climate and hydrology. *Nat. Commun.* **5**, 1–8 (2014).
- Liang, C., Schimmel, J. P. & Jastrow, J. D. The importance of anabolism in microbial control over soil carbon storage. *Nat. Microbiol.* **2**, 17105 (2017).
- Kind, T. & Fiehn, O. Advances in structure elucidation of small molecules using mass spectrometry. *Bioanal. Rev.* **2**, 23–60 (2010).
- Broeckling, C. D., Afsar, F. A., Neumann, S., Ben-Hur, A. & Prenni, J. E. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal. Chem.* **86**, 6812–6817 (2014).
- Dunn, W. B. et al. Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics* **9**, 44–66 (2013).
- Feunang, Y. D. et al. ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *J. Cheminform.* **8**, 1–20 (2016).
- Allen, F., Pon, A., Wilson, M., Greiner, R. & Wishart, D. CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra. *Nucleic Acids Res.* **42**, 94–99 (2014).
- Xia, J. & Wishart, D. S. Using MetaboAnalyst 3.0 for comprehensive metabolomics data analysis. *Curr. Protoc. Bioinformatics* **55**, 10.1–14.10.91 (2016).
- Szymańska, E., Saccenti, E., Smilde, A. K. & Westerhuis, J. A. Double-check validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* **8**, 3–16 (2012).
- Oksanen, J., Kindt, R., Legendre, P., O'Hara, B. & Stevens, H. H. The vegan package. *Community Ecol. Package* **10**, 631–637 (2007).
- Feggl, T. S., Baron, J. S., Fountain, A. G., Johnson, G. F. & Hall, E. K. The differing biogeochemical signatures of glaciers and rock glaciers. *J. Geophys. Res. Biogeosci.* **121**, 919–932 (2016).
- Chen, W., Westerhoff, P., Leenheer, J. A. & Booksh, K. Fluorescence excitation - emission matrix regional integration to quantify spectra for dissolved organic matter. *Environ. Sci. Technol.* **37**, 5701–5710 (2003).
- Creed, I. I. F. et al. The river as a chemostat: fresh perspectives on dissolved organic matter flowing down the river continuum. *Can. J. Fish Aquat. Sci.* **14**, 1–14 (2015).
- Burns, M. A., Barnard, H. R., Gabor, R. S., McKnight, D. M. & Brooks, P. D. Dissolved organic matter transport reflects hillslope to stream connectivity during snowmelt in a montane catchment. *Water Resour. Res.* **52**, 4905–4923 (2016).
- Ward, N. D. et al. Where carbon goes when water flows: carbon cycling across the aquatic continuum. *Front. Mar. Sci.* **4**, 1–27 (2017).
- Schimel, J., Balsler, T. C. & Wallenstein, M. Microbial stress-response physiology and its implications for ecosystem function. *Ecology* **88**, 1386–1394 (2007).
- Kaiser, K. & Kalbitz, K. Cycling downwards—dissolved organic matter in soils. *Soil Biol. Biochem.* **52**, 29–32 (2012).
- Hättenschwiler, S. & Vitousek, P. M. The role of polyphenols in terrestrial ecosystems nutrient cycling. *Tree* **15**, 238–243 (2000).
- Rumpel, C. & Kögel-Knabner, I. Deep soil organic matter - a key but poorly understood component of terrestrial C cycle. *Plant Soil* **338**, 143–158 (2011).
- Raymond, P. A., Saiers, J. E. & Sobczak, W. V. Hydrological and biogeochemical controls on watershed dissolved organic matter transport: pulse- shunt concept. *Ecology* **97**, 5–16 (2016).
- Marcatelli, A., Baxter, C. V., Mineau, M. M. & Hall, R. O. Quantity and quality: unifying food web and ecosystem perspectives on the role of resource subsidies in freshwaters. *Ecology* **92**, 1215–1225 (2011).
- Torres-Ruiz, M., Wehr, J. D. & Perrone, A. A. Trophic relations in a stream food web: importance of fatty acids for macroinvertebrate consumers. *J. North Am. Benthol. Soc.* **26**, 509–522 (2007).
- Boye, K. et al. Thermodynamically controlled preservation of organic carbon in floodplains. *Nat. Geosci.* **10**, 415–420 (2017).
- Danczak, R. E. et al. Seasonal hyporheic dynamics control coupled microbiology and geochemistry in Colorado River sediments. *J. Geophys. Res. Biogeosci.* **121**, 2976–2987 (2016).
- Cory, R. M., Harrold, K. H., Neilson, B. T. & Kling, G. W. Controls on dissolved organic matter (DOM) degradation in a headwater stream: the influence of photochemical and hydrological conditions in determining light-limitation or substrate-limitation of photo-degradation. *Biogeosciences* **12**, 6669–6685 (2015).
- Opsahl, S. P. & Zepp, R. G. Photochemically-induced alteration of stable carbon isotope ratios ( $\delta C-13$ ) in terrigenous dissolved organic carbon. *Geophys. Res. Lett.* **28**, 2417–2420 (2001).
- Brooks, P. D. et al. Hydrological partitioning in the critical zone: Recent advances and opportunities for developing transferable understanding of water cycle dynamics. *Water Resour. Res.* **51**, 6973–6987 (2015).
- Wetzel, R. G., Hatcher, P. G., Bianchi, T. S. & Dec, N. Natural photolysis by ultraviolet irradiance of recalcitrant dissolved organic matter to simple substrates for rapid bacterial metabolism. *Limnol. Oceanogr.* **40**, 1369–1380 (2007).
- Moran, M. A. & Zepp, R. G. Role of photoreactions in the formation of biologically labile compounds from dissolved organic matter. *Limnol. Oceanogr.* **42**, 1307–1316 (1997).
- Cotrufo, M. F., Wallenstein, M. D., Boot, C. M., Denef, K. & Paul, E. A. The Microbial Efficiency-Matrix Stabilization (MEMS) framework integrates plant litter decomposition with soil organic matter stabilization: do labile plant inputs form stable soil organic matter? *Glob. Chang. Biol.* **19**, 988–995 (2013).
- Halket, J. M. et al. Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *J. Exp. Bot.* **56**, 219–243 (2005).
- Williams, D. H., Stone, M. J., Hauck, P. R. & Rahman, S. K. Why are secondary metabolites (natural products) biosynthesized? *J. Nat. Prod.* **52**, 1189–1208 (1989).
- Horner, J. D., Gosz, J. R. & Cates, R. G. The role of carbon-based plant secondary metabolites in decomposition in terrestrial ecosystems. *Am. Nat.* **132**, 869–883 (1988).
- Challis, G. L. & Hopwood, D. A. Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces* species. *Proc. Natl Acad. Sci. USA* **100**, 14555–14561 (2003).
- Covino, T., McGlynn, B. & McNamara, R. Land use/land cover and scale influences on in-stream nitrogen uptake kinetics. *J. Geophys. Res. Biogeosci.* **117**, 1–13 (2012).
- United States Environmental Protection Agency. *Connectivity of Streams and Wetlands to Downstream Waters: A Review and Synthesis of the Scientific Evidence*. Report No. EPA/600/R-14/475F (U.S. Environmental Protection Agency, Washington, DC, 2015).

54. Baron, J. S. et al. Ecosystem responses to nitrogen deposition in the Colorado Front Range. *Ecosystems* **3**, 352–368 (2000).
55. Kendall, K. A., Shanley, J. B. & McDonnell, J. A hydrometric and geochemical approach to test the transmissivity feedback hypothesis during snowmelt. *J. Hydrol.* **219**, 188–205 (1999).
56. Trussell, R. & MD, U. The formation of trihalomethanes. *Am. Water Works Assoc. J.* **70**, 604–612 (1978).
57. Maas, A., Dozier, A., Manning, D. T. & Goemans, C. Water storage in a changing environment: the impact of allocation institutions on value. *Water Resour. Res.* **53**, 672–687 (2017).
58. Puttock, A., Graham, H. A., Cunliffe, A. M., Elliott, M. & Brazier, R. E. Eurasian beaver activity increases water storage, attenuates flow and mitigates diffuse pollution from intensively-managed grasslands. *Sci. Total Environ.* **576**, 430–443 (2017).
59. Barnett, T. P. et al. Human-induced changes in the hydrology of the Western United States. *Science* **319**, 1080–1084 (2008).
60. Braddock, W. & Cole, J. C. *Geologic Map of Rocky Mountain National Park and Vicinity: Miscellaneous Investigations Series Map I-1973, Scale 1:50,000* (U.S. Geological Survey, Lakewood, 1990).
61. Kramer, N., Wohl, E. E. & Harry, D. L. Using ground penetrating radar to ‘unearth’ buried beaver dams. *Geology* **40**, 43–46 (2012).
62. Fellman, J. B., Hood, E. & Spencer, R. G. M. Fluorescence spectroscopy opens new windows into dissolved organic matter dynamics in freshwater ecosystems: a review. *Limnol. Oceanogr.* **55**, 2452–2462 (2010).
63. Cory, R. M. & McKnight, D. M. Fluorescence spectroscopy reveals ubiquitous presence of oxidized and reduced quinones in dissolved organic matter. *Environ. Sci. Technol.* **39**, 8142–8149 (2005).
64. Hothorn, T. et al. lmtest: testing linear regression models. R Package Version 0.9-3.4 (R Foundation for Statistical Computing, Vienna, 2015).
65. R Core. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, 2014).
66. Lenth, R. Package lsmeans: least-squares means. *CRAN Repos* (2016).
67. van den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K. & van der Werf, M. J. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* **7**, 142 (2006).
68. Grootveld, M. in *Metabolic Profiling: Disease and Xenobiotics* 1–34 (Royal Society of Chemistry, Cambridge, 2014).
69. Wehrens, R. & Mevik, B.-H. pls: partial least squares regression (PLSR) and principal component regression (PCR). *R Package Version 2.1-0* (2007).
70. Woolf, D., Lehmann, J. & Lee, D. R. Optimal bioenergy power generation for climate change mitigation with or without carbon sequestration. *Nat. Commun.* **7**, 1–11 (2016).

## Acknowledgements

This research was supported by a National Science Foundation CAREER grant (number 255228) awarded to M.D.W., and a National Science Foundation IGERT fellowship (number 0966346) awarded to L.M.L. and N.A.S. We thank two anonymous reviewers for their thoughtful comments, which significantly improved our manuscript. We thank Dr. Ellen Wohl for assistance in selecting study sites, Pam Wegener for sharing discharge data, and Dr. Ann Hess for statistics consultation.

## Author contributions

L.M.L., N.A.S., and T.S.F. designed the study, with input from C.M.B., T.P.C., and M.D.W.; L.M.L., N.A.S., and T.S.F. conducted field work and analyzed the data; L.M.L. wrote the first draft of the manuscript; N.A.S., T.P.C., C.M.B., T.S.F., and M.D.W. commented and contributed to manuscript revisions.

## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-019-08406-8>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Journal peer review information:** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019