

## Linking stream microbial community functional genes to dissolved organic matter and inorganic nutrients

Christina Fasching <sup>1\*,a</sup> Christian Akotoye,<sup>1</sup> Mina Bižić,<sup>2</sup> Jeremy Fonvielle <sup>2,b</sup> Danny Ionescu,<sup>2</sup> Sabateeshan Mathavarajah,<sup>1</sup> Luca Zoccarato,<sup>2</sup> David A. Walsh <sup>3</sup> Hans-Peter Grossart <sup>2,4\*</sup> Marguerite A. Xenopoulos <sup>1</sup>

<sup>1</sup>Department of Biology, Trent University, Peterborough, Ontario, Canada

<sup>2</sup>Leibniz Institute for Freshwater Ecology and Inland Fisheries, Stechlin, Germany

<sup>3</sup>Department of Biology, Concordia University, Montreal, Quebec, Canada

<sup>4</sup>Institute for Biochemistry and Biology, Potsdam University, Potsdam, Germany

### Abstract

There is now increasing evidence for the importance of microbial regulation of biogeochemical cycling in streams. Resource availability shapes microbial community structure, but less is known about how landscape-mediated availability of nutrients and carbon can control microbial functions in streams. Using comparative metagenomics, we examined the relationship between microbial functional genes and composition of dissolved organic matter (DOM), nutrients, and suspended microbial communities in 11 streams, divided into three groups based on the predominant land cover category (agriculture, forested, or wetland). Using weighted gene co-occurrence network analysis, we identified clusters of functions related to DOM composition, agricultural land use, and/or wetland and forest land cover. Wetland-dominated streams were characterized by functions related to nitrogen metabolism and processing of aromatic carbon compounds, with strong positive correlations with dissolved organic carbon concentration and DOM aromaticity. Forested streams were characterized by metabolic functions related to monomer uptake and carbohydrates, such as mannose and fructose metabolism. In agricultural streams, microbial functions were correlated with more labile, protein-like DOM, PO<sub>4</sub>, and NO<sub>3</sub>, likely reflecting functional adaptation to labile DOM and higher nutrient concentrations. Distinct changes in the functional composition and loss of functional diversity of microorganisms became evident when comparing natural to agricultural catchments. Although all streams showed signs of functional redundancy, loss of species richness per function in agricultural catchments suggests that microbial functions in natural catchments may be more resilient to disturbance. Our results provide new insight into microbial community functions involved in nutrient and carbon biogeochemical cycles and their dependence on specific environmental settings.

Microbes are essential for regulating biogeochemical systems such as the mineralization of organic matter and nutrient uptake and transformation (Battin et al. 2003; Zhao et al. 2014; Vachon

et al. 2017). In stream ecosystems, microbial communities characterized by their taxonomic composition and function are essential to carry out processes in nutrient and carbon cycling. Stream microbial communities are affected by stream water chemistry which in turn is influenced by these same microbes and their metabolic pathways. A stream's chemical environment is also profoundly influenced by the land use and land cover of the surrounding landscape. A stream's land use will dictate its nutrient concentration, dissolved organic matter (DOM) quantity and quality, water temperature, and hydrology (Wilson and Xenopoulos 2008; Williams et al. 2010, 2016; Wilhelm et al. 2015) and, consequently, the metabolic functions of its microbial communities (Williams et al. 2010; Comte et al. 2013a; Wilhelm et al. 2013; Peter and Sommaruga 2016). In light of changing climate and land use, the links between stream biogeochemistry, microbial community composition, and metabolic functions need to be better understood (Findlay 2010).

\*Correspondence: christina.fasching@ufz.de; hgrossart@igb-berlin.de

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Additional Supporting Information may be found in the online version of this article.

<sup>a</sup>Present address: Department of Soil Ecology, Helmholtz Centre for Environmental Research – UFZ, Halle (Saale), Germany

<sup>b</sup>Present address: Department of Plant Sciences, University of Cambridge, Cambridge, United Kingdom

**Special Issue:** Linking Metagenomics to Aquatic Microbial Ecology and Biogeochemical Cycles. Edited by: Hans-Peter Grossart, Ramon Massana, Katherine McMahon and David A. Walsh.

The processing of nitrogen, phosphorus, and DOM are functions that have been linked to specific microbial families (Wilhelm et al. 2015; Traving et al. 2017; LeBrun et al. 2018). For example, *Cyanobacteria*, *Proteobacteria*, and some *Actinobacteria* are key players in the cycling of nitrogen in freshwater ecosystems (Newton et al. 2011; Osman et al. 2017) because of their ability to fix available CO<sub>2</sub> and N<sub>2</sub> via the nitrogenase enzyme complex (Gtari et al. 2012; Delmont et al. 2018). However, phylogeny and microbial functions are not always directly linked as functional redundancy on a species level or functional similarity of communities may muddle our understanding of the relationships between microbial community structure and functional/metabolic responses (Allison and Martiny 2008; Comte et al. 2013a).

In order to better understand ecosystem function and how organisms respond to environmental change, microbial traits and function must be considered together with taxonomy (Green et al. 2008; Dopheide et al. 2015; Osterholz et al. 2016). Indeed, the dynamics of indicator genes involved in major pathways of carbon and nutrient cycling (e.g., in the case of nitrogen: nitrogen fixation, ammonia oxidation, nitrate and nitrite reduction) depend on both species functional composition and environmental conditions (Dopheide et al. 2015). Nevertheless, the relationship between taxonomy and function is not always clear and studies have shown that there is a complex mix of influences. Some studies point to community history or taxonomy as important drivers of function (Langenheder et al. 2006). On the other hand, environmental gradients have been shown to be more important in driving functional composition of a community (Ruiz-González et al. 2015). As an example, Ruiz-González et al. (2015) found that a DOM quality gradient is correlated to functional microbial traits but not the taxonomic profile of the stream microbial community. As another example, urbanization and the resulting changes in DOM quality (Williams et al. 2016) have been linked to taxonomic composition of freshwater ecosystems resulting in the loss of keystone taxa (Hosen et al. 2017) and changes in community structure and metabolism (Medeiros et al. 2014). These examples underscore current knowledge gaps with respect to trait distribution and functional redundancy, and how they are linked to the microbial taxonomic composition within their environment. Understanding the relationship between microbial community structure and functional specialization in response to resources may shed new light on carbon and nutrients cycling (Battin et al. 2008). These relationships between microbial structure and function could then be used to better understand land use change on biogeochemical processes.

Despite growing interest in uncovering how functional and/or taxonomic traits relate to environmental conditions, there is still a lack of research examining microbial genes and functions in lotic ecosystems, especially across environmental gradients and when compared to marine and lentic ecosystems (Zeglin 2015) and also in relation to DOM quality. Here, we used a metagenomics approach to assess the distribution of microbial

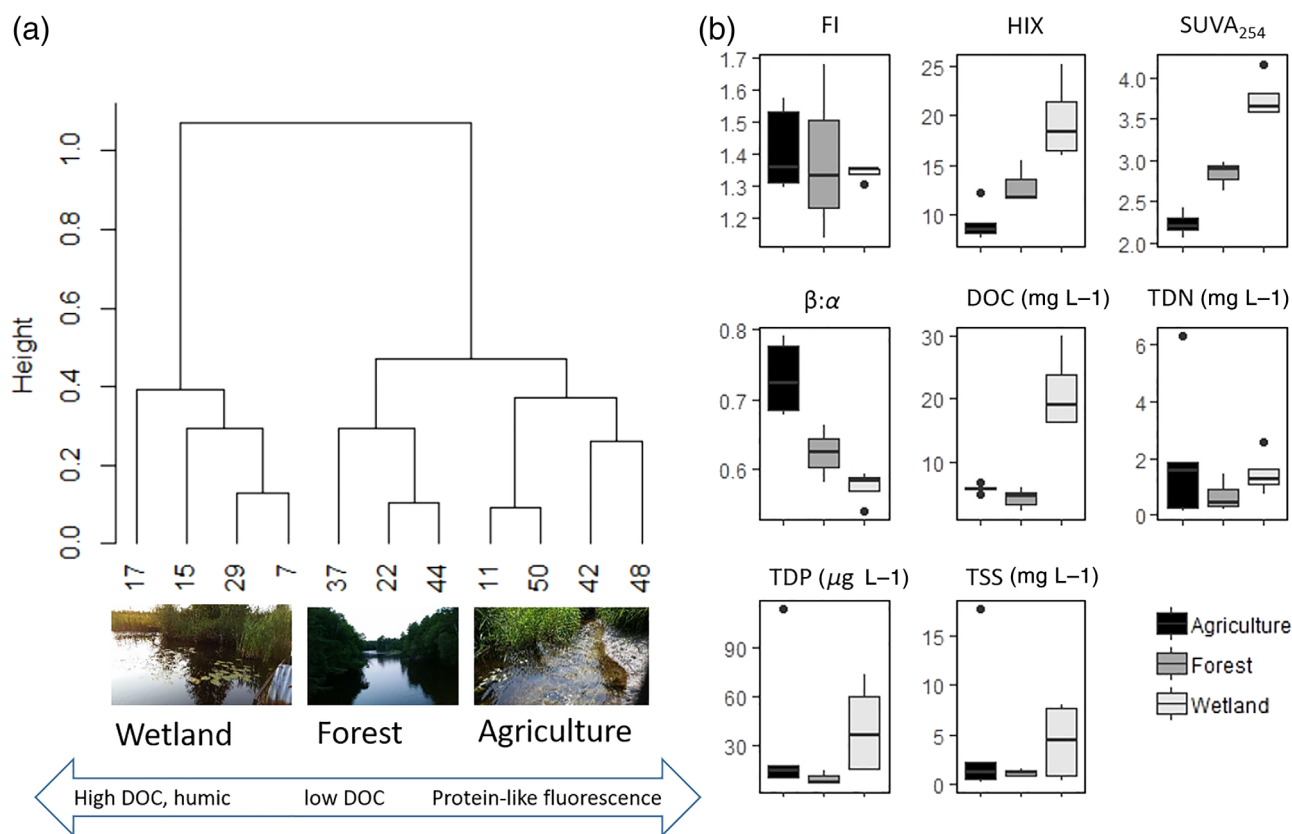
functions and metabolic pathways as well as taxonomic groups across streams differing in their land use and land cover composition. We investigated how functions are related to the stream's physico-chemical environment, how functional diversity changes in response to nutrients and DOM composition, and whether functional similarity between communities is present. We anticipated metabolic plasticity across streams and expected to observe differences in community functional profile with respect to changes in catchment land use and subsequent changes in nutrient concentration and DOM composition. We hypothesized that agricultural streams have a greater proportion of genes involved in nitrogen, phosphorus, and organic matter transformation due to the greater input of nutrients and distinct DOM quality compared to natural land cover (Wilson and Xenopoulos 2009; Williams et al. 2016; Fuß et al. 2017). We identified clusters of functions related to specific DOM quality, directly driven by the landscape. Through this approach, we can better understand and evaluate the relationship between land use and microbial functions.

## Methods

### Site characterization

We analyzed the microbial community taxonomic composition, functional genes, and DOM and nutrient composition in 11 streams located across Southern Ontario, Canada (Supporting Information Fig. S1). Catchment sizes ranged from 18.1 to 354.3 km<sup>2</sup> with the surface geology composed of a mixture of clay, gravel, sand, and organic deposited soils. Land cover classifications for the entire watershed and within a 100-m buffer along either side upstream of the sampling site (riparian zone) have been previously determined (D'Amario and Xenopoulos 2015). DOM composition and nutrients in these streams are strongly correlated with catchment-scale attributes (Wilson and Xenopoulos 2008, 2009; Williams et al. 2010).

Streams were grouped based on the type of land cover, nutrient concentrations, and DOM composition using a cluster analysis (*hclust* function in *base* R package; R Core Team 2017) and the robustness of clusters was tested using the analysis of similarity (*anosim* function in the *Vegan* package (Oksanen et al. 2017)). The streams were separated into three categories (agriculture, forest, and wetland; one-way ANOSIM test,  $R = 0.34$ ,  $p < 0.05$ ) based on whether their catchment was predominantly covered by agriculture, forested land cover, or wetland land cover. This separation was further determined by the stream's respective nutrient and carbon concentrations with agriculture and wetland sites having the highest total dissolved phosphorus (TDP) and total dissolved nitrogen (TDN) concentrations and forested sites having the lowest nutrient and dissolved organic carbon (DOC) concentrations. DOM composition was also important in clustering the sites with wetland and forested sites having more aromatic DOM and agricultural sites having more protein-like DOM (Fig. 1).



**Fig. 1.** (a) Cluster analysis of the DOM and nutrient composition of all sites ( $n = 11$ ). The height represents Euclidian distance. (b) Box-whisker plots of DOM composition, DOC, TSS, TDN, and TDP per land cover category as defined by the cluster analysis ( $n = 3-4$ ).

### DNA extraction and sequencing

Samples for DNA extraction were collected 10 cm below the water's surface at the thalweg of each stream in May 2016. For each stream, approximately 5 L of water were collected and transported on ice to the lab for further processing. In the lab, metazoans and larger particles were removed from the samples using 0.7  $\mu\text{m}$  GF/F filters and the remaining planktonic microbial communities were collected on a 0.2  $\mu\text{m}$  membrane filter (250–500 mL of water was filtered). Although no large colonial cyanobacteria were visibly present at the time of sampling, pre-filtering using 0.7  $\mu\text{m}$  pore size may have excluded some of the larger bacteria and cyanobacteria from further analysis. Filters with the microbial community ( $>0.2 \mu\text{m}$  but  $<0.7 \mu\text{m}$ ) were stored in 2 mL of RNAlater (Nishimoto et al. 2007) and kept at  $-20^\circ\text{C}$  until processing. Within 4 weeks, DNA extraction was performed following a protocol adapted from Zhou et al. (1996). Briefly, cells were pelleted from the RNAlater solution and lysed using Phenol-Chloroform-Isoamylalcohol (PCI), 20% sodium dodecyl sulfate (SDS), and 0.7 mm zirconium beads. This solution was then heat shocked for 10 min at  $60^\circ\text{C}$ . Repeated centrifugation and PCI treatment were used to completely remove cellular debris. DNA was then precipitated with 96% ethanol, quantified through fluorometric quantification using the Qubit 2.0 fluorometer and subsequently stored in TE buffer at  $-80^\circ\text{C}$  till further analysis. Polymerase chain reaction (PCR) reactions for

the 16S rRNA gene were performed to confirm the presence microbial DNA in these samples.

After extraction, the genetic material was stored in TE buffer at  $-80^\circ\text{C}$  for 2 weeks. Following DNA extraction, samples were sent for Illumina Sequencing (HiSeq 2X150 bp, MrDNA Lab, Texas, U.S.A.). The libraries were prepared using Nextera DNA Sample preparation kit (Illumina) following the manufacturer's user guide. The libraries were pooled and diluted (to 10 pmol  $\text{L}^{-1}$ ) and sequenced paired end for 300 cycles using the HiSeq system (Illumina). Low quality or short reads ( $Q < 15$ , length  $< 100$  bp), adapter sequences, and PhiX contaminations were removed using the BBduk tool of the BBMap program package (Bushnell 2014). The data were analyzed within the MG-RAST pipeline (Keegan et al. 2016) using the default parameters including quality and adapter trimming, removal of contaminant PhiX and human sequences, identification and annotation of rRNA sequences, and functional and taxonomic annotation of protein coding sequences. Annotations with an E-value higher than  $10^{-5}$  were discarded. The total number of sequences submitted, and the percent of those that were annotated are presented in Supporting Information Table S2. Overall, more than 90% of the sequences in each sample passed the MG-RAST quality control and of these 80–96% were predicted have a function. Of the latter 36–56% (1.2–5.4 million sequences) could be assigned to known functions or ribosomal RNA. We used the SEED database

(Overbeek et al. 2005) and generated a total number of assigned functions of 7585 of which 5526 were used for further analyses, after unclassified reads and eukaryotes were excluded.

Individual assemblies generated using Spades v. 3.06 (Bankevich et al. 2012) were initially uploaded to the Integrated Microbial Genomes (IMG) annotation server (Chen et al. 2016). These assemblies while likely generated reliable information due to the increased sequence length as compared to the raw data, nevertheless provided a shallow overview of the community structure and composition of each sample due to insufficient coverage of all taxa in such diverse environments. Therefore, we chose to use the MG-RAST annotation which better encompasses the full extent of the data. Taxonomic information was derived in two manners. First, we used 16S rRNA genes as detected in the metagenomic data for an accurate taxonomic identification of the dominant microbial families in our samples. Since rRNA genes are as abundant as any other gene in DNA-based data, these do not describe the entire extent of the microbial community as evidenced also by the low number (32) of families identified. As such, to assign taxonomic identities to proteins used to determine the degree of functional redundancy, the RefSeq database (implemented in MG-RAST, <https://www.ncbi.nlm.nih.gov/pubmed/17130148>) was used. While using this database generated significantly more taxonomic information (233) families, these data may be biased toward organisms for which genomic data is available or for which specific proteins have been studied. The data are available the European Nucleotide Archive (ENA; Project number: PRJEB34059) (<https://www.ebi.ac.uk/ena>).

### Water quality and environmental variables

Specific conductivity, salinity, and temperature were directly measured at each site using a handheld probe (YSI probePro30, Yellow Springs, Ohio, U.S.A.). Dissolved oxygen (DO) was measured using a separate probe (Hach HQ30D, Loveland, Colorado). Flow velocity was measured at both the riffle and pool area of each stream using a flow meter (Marsh-McBirney Flo-Mate 2000, Maryland, U.S.A.). Discharge was monitored using an acoustic Doppler velocimeter (SonTek FlowTracker, San Diego, California, U.S.A.). For larger streams, flow data were obtained from the closest Water Survey of Canada hydrological metering station on that stream. Samples for TDP and TDN were collected at each site in acid washed 50 mL Falcon tubes. Samples for dissolved nutrients were collected by passing the water first through a 0.7  $\mu\text{m}$  GF/F filter and then, through 0.2  $\mu\text{m}$  polycarbonate filters. Water samples for DOC concentration and composition (DOM) analysis were sterile-filtered (prerinsed 0.2  $\mu\text{m}$  membrane filters: Whatman, Mississauga, Ontario, Canada) and stored in amber bottles (acid-rinsed and combusted at 450°C, 4 h) at 4°C until analysis within 4 weeks. TDP concentrations were determined by spectrophotometry (Varian Cary 50) using the molybdate colorimetry method (APHA 1992). TDN and  $\text{NO}_3^-$  were digested using persulfate prior to measurements by spectrophotometry (Cary-50, Varian, Palo Alto, California, U.S.A.). DOC was measured by heated persulfate wet oxidation using an OI Aurora TOC analyzer

(College Station, Texas, U.S.A.). To remove inorganic carbon, each sample was acidified using hydrochloric acid and purged with ultra-pure oxygen (Wilson and Xenopoulos 2008). Total suspended solids (TSS) were determined by filtering 350 mL of sample through precombusted GF/F filters, which were then dried at 60°C, stored at 20°C, and weighed using a microbalance.

### DOM absorbance and fluorescence

For each DOM sample, absorbance spectra were generated from 800 to 200 nm with an increment of 5 nm using a Cary 50 spectrophotometer (Agilent Technologies, Mississauga, ON, Canada). Absorbance at 254 nm was divided by the DOC concentration ( $\text{mg C L}^{-1}$ ) to determine the specific UV absorbance ( $\text{SUVA}_{254}$ ). The  $\text{SUVA}_{254}$  was used as an indicator of DOM aromaticity (Weishaar et al. 2003). The molar absorptivity ( $\epsilon_{280}$ ), measured as the specific UV absorbance at 280 nm, was used as an additional indicator of aromaticity (Weishaar et al. 2003). Using fluorescence spectroscopy (Varian Cary Eclipse Fluorescence Spectrophotometer, Agilent), excitation-emission matrices (EEMs) were generated by scanning fluorescence intensity over a range of excitation (230–500 nm at 5-nm increments) and emission wavelengths (270–600 nm at 2-nm increments). EEMs were corrected for the inner filter effect and second-order Raman and Rayleigh scatter effects. All samples were blank-corrected by subtracting the EEM of MilliQ water (Millipore) and subsequently converted to Raman units (RU). Additional fluorescence indices were calculated as well. The humification index (HIX) is directly proportional to the humic content of DOM, with higher values indicating more humic-like DOM (Zsolnay et al. 1999). The fluorescence index is inversely related to the lignin content of DOM, in which lower values ( $\sim 1.3$ ) suggest a more terrestrially derived DOM source whereas higher values ( $\sim 1.8$ ) suggest microbial-derived DOM (McKnight et al. 2001). We measured the freshness index ( $\beta/\alpha$ ) which indicates recent production of DOM, with  $\beta$  representing more recently derived DOM and  $\alpha$  representing relatively more decomposed DOM (Wilson and Xenopoulos 2008). Furthermore, we fitted the EEMs to the existing parallel factor analysis (PARAFAC) (Stedmon and Bro 2008) model developed by (Williams et al. 2013) using the DOMFluor Toolbox (1.7; containing the N-Way toolbox, 3.1) (Andersson and Bro 2000) and expressed the intensity of the components in relative terms (% total fluorescence). Briefly, the model extracted seven components which are C1: ubiquitous humic-like, C2 and C3: terrestrial humic-like, C4: soil fulvic-like, C5: microbial humic-like, C6: anthropogenic microbial humic-like, and C7: protein-like component (Supporting Information Fig. S2).

### Statistical analysis

MG-RAST annotated gene counts were normalized to the total number of genes per metagenome to account for the variation in metagenome sizes across streams. These relative gene abundances were then used to identify gene functions related to the environmental data. We calculated functional diversity using Shannon H as a diversity index. Furthermore, we counted the

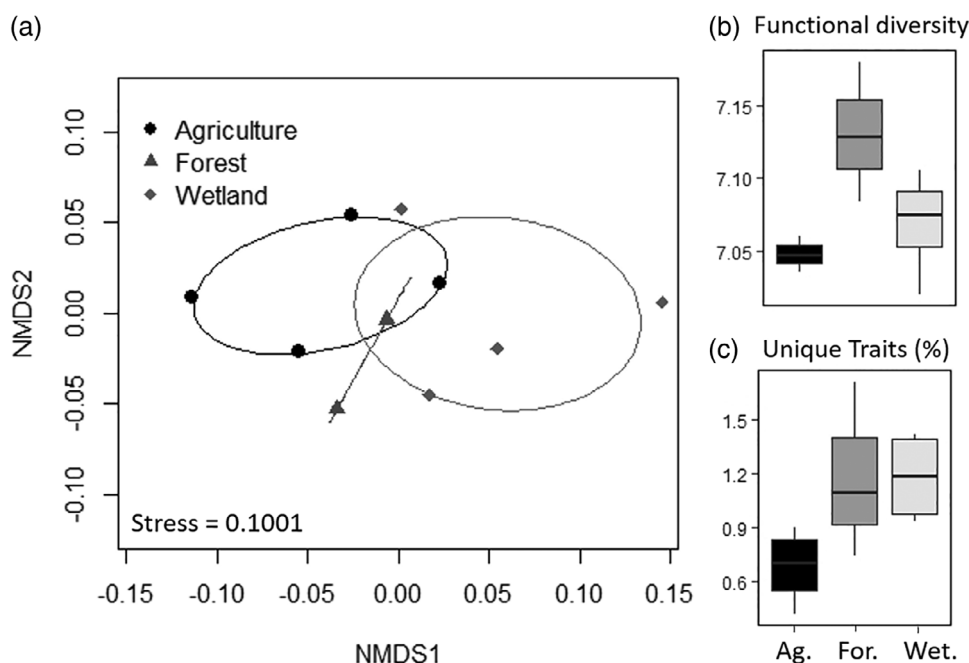
number of unique functions per agriculture, forest, or wetland category (functions that occur exclusively in each land use category). Our taxonomic analysis was restricted here as the taxonomy data were derived from the functional dataset through the MG-RAST pipeline (Keegan et al. 2016). However, this unique data set allows the examination of functional diversity by comparing our taxonomic composition per function across sites.

Nonmetric multidimensional scaling analysis based on Bray–Curtis dissimilarities was applied to illustrate differences in microbial functional composition across land cover categories. To test for differences in the functional composition per land cover category, we completed an ANOVA (Permutational multivariate analysis of variance (PERMANOVA), using the Adonis function in R Vegan package; Oksanen et al. 2017) using distance Matrices (Bray–Curtis dissimilarity). All the statistical analyses were performed using the statistical environment R (R Core Team 2017) and the packages Vegan (Oksanen et al. 2013) and weighted gene co-occurrence network analysis (WGCNA) (Langfelder and Horvath 2008).

To identify microbial community functions that were related to DOM composition and nutrients, we used WGCNA (Langfelder and Horvath 2008). Briefly, genes were clustered into gene modules (=genecluster) via unsigned co-occurrence network (minimum of 20 genes per module). Automatic block-wise module detection preclusters genes into large clusters, referred to as blocks, using a variant of k-means clustering. Gene modules are then defined for each block by means of hierarchical clustering and the modules with highly correlated ( $r > 0.75$ ) eigengenes are eventually merged. These gene modules were summarized by their first principal component

(eigengene) and correlated to the DOM composition, PARAFAC components, and various nutrients (TDN, TDP, DO, and DOC). Correlations and  $p$  values were obtained from a univariate regression model between each module eigengene and each environmental trait. Gene significance (GS) was calculated for each gene relative to a specific environmental variable by taking the log of the  $p$  value. Nonzero values (positive or negative) of GS indicated that the respective gene was well related to the specific environmental factor. Last, module significance is determined as the average absolute GS measure for all genes in a module (Langfelder and Horvath 2008).

Ternary plots were used to visualize how the functional community composition varied across the sites based on land use and DOM composition. Although all functions (5526) are shown, we focused on the 20 modules (containing 2343 functions) which were correlated ( $r > 0.62$ ) with DOM composition and/or nutrient concentrations. Points close the corners of the plot represent functions/taxa that are abundant in the respective land cover category. Microbial families ( $n = 32$ ) and functions ( $n = 2343$ ) were classified as “specialized” for each forested, wetland, or agricultural category if their relative abundance within that category was greater than 60% (indicated by the solid line within the plots), as an arbitrary threshold (Wilhelm et al. 2015). All others were classified as generalists (taxa) or functionally conserved (gene modules). Using this threshold, we assigned 2086 generalist functions, and 549 specialist functions. We used a tree map to visualize the most abundant functions per module, limiting the dataset to major pathways related to carbon, nitrogen, and phosphorus cycling (374 functions). This was done in order to identify “indicator



**Fig. 2.** (a) Non-metric multidimensional scaling (NMDS) of all functions ( $n = 5526$ ) across land cover categories (agriculture, forest, and wetland) ( $n = 11$ ). (b) Functional diversity as the Shannon index [H] and (c) unique traits in percent across the landscape categories ( $n = 3-4$ ).

functions” for biogeochemical cycling in the respective land cover category.

While the ternary plots relate agriculture, forest, and wetlands to the functions via the DOM and nutrient gradient, we also used an indicator species analysis to directly identify indicator traits for each land cover category. To identify phylogenetically coherent microbial traits within our three land cover categories, we used the LefSE method (Linear discriminant analysis Effect Size), (Segata et al. 2011). Nonparametric factorial Kruskal–Wallis sum rank tests ( $\alpha = 0.05$ ) were used to detect different abundant features (i.e., subsystems, functions) in each agriculture, forest, and wetland categories. The phylogenetic consistency was tested using pairwise Wilcoxon rank-sum tests ( $\alpha = 0.05$ ) and estimates of effect size of each abundant feature using linear discriminant analysis. All classes were compared (most stringent) and a value of 2.0 of the logarithmic linear discriminant analysis score was chosen as a threshold for discriminative features.

To investigate potential functional redundancy within our sites, we visualized the microbial taxonomic composition related to the individual functions identified as “specialized.” Only families with a relative abundance above 5% per function were visualized. We determined whether the same functions were related to the same set of microbial taxa when the land cover category changed or, alternatively, whether the function was found to be related to a different community. In short, we were asking whether the functional composition depends on the taxonomical community composition or whether the same functions can be carried out by different communities in other land cover/land use categories. To further examine whether function could be linked to a large range of microbial families, we computed alpha diversity of the remaining microbial taxonomic composition per function using the inversed Simpson index using the R Vegan package (Oksanen et al. 2017). To test for the similarity in the taxonomic composition of different functions, we computed a Principal Coordinates Analysis (PCoA) based on the Bray–Curtis dissimilarities per “specialist function” per land cover category using the R Vegan package (Oksanen et al. 2017) (A PCoA of all functions can be found in the Supporting Information). We also tested the species richness related to the individual functions and land cover category using corrected (Sokol and Rohlf 1995) paired *t*-tests in order to better understand the potential resilience of functions to disturbance and species loss. High species richness per function would indicate an increased resilience of functions to disturbance.

## Results

### DOM and nutrient composition

Across the sample sites, TDP and TDN varied from 7.56 to 78.2  $\mu\text{g L}^{-1}$  and 0.15 to 2.56  $\text{mg L}^{-1}$ , respectively (Fig. 1). TDN, TDP, and DOC concentrations were elevated in wetland-dominated streams compared to forested sites (Fig. 1). DOC and DO concentrations were also highly variable and ranged from 2.11 to 30.0  $\text{mg L}^{-1}$  and 3.89 to 15.6  $\text{mg L}^{-1}$ , respectively. DOM

**Table 1.** Catchment land cover, functional diversity, evenness, richness, and number of unique functions per site. Numbers indicate land use in the riparian zones of the individual catchments.

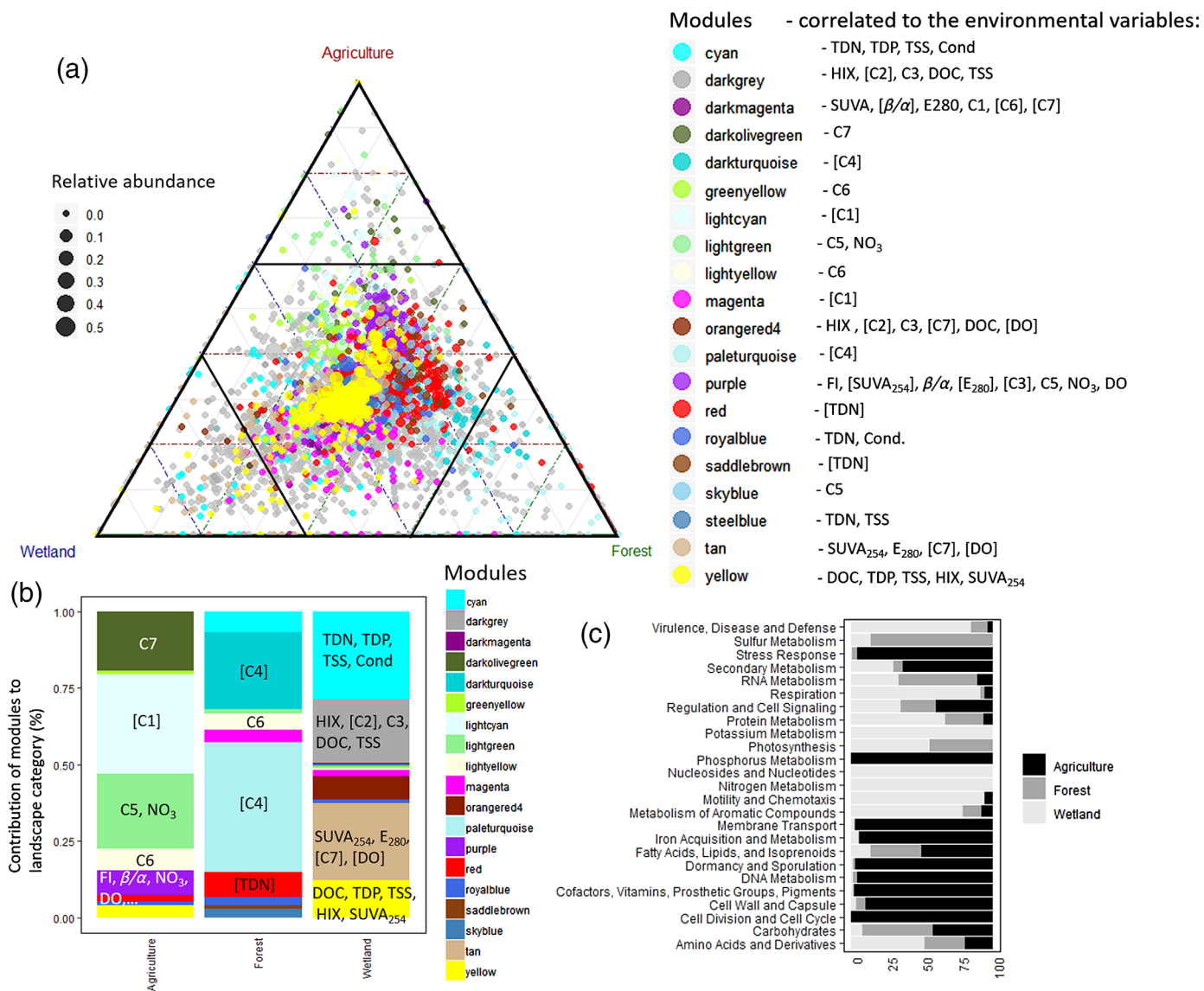
Site	Stream	Description	Total cropland (%)	Wooded area (%)	Wetland (%)	Urban areas (%)	Rural land use (%)	Functional diversity	Functions evenness	Functions richness	Unique functions (%)	Unique species (%)
11	Talbot River	Agriculture	31.3	12.3	34.9	1	9.2	7.04	2.09	4168	0.6	2.45
42	Schomber Canal	Agriculture	72.3	4.5	3	2.5	13.3	7.04	1.96	4027	0.42	1.23
48	Fish Creek	Agriculture	49.2	11.8	0.1	0.5	32.9	7.06	1.81	4407	0.9	1.23
50	Indian River	Agriculture	15	48.6	10.9	1.5	4.2	7.05	1.95	4337	0.81	5.52
22	Burnt River	Forest	3.1	59.1	26.1	0	4.9	7.13	1.78	4050	0.74	7.98
37	Baxter Creek	Forest	16.5	40	10.3	3.4	9.3	7.08	1.83	4582	1.1	2.45
44	Coldwater Creek	Forest	1.2	50.3	32.6	2.7	2.1	7.18	1.9	4662	1.71	3.07
7	Mariposa Brook	Wetland	7.4	11.1	58.7	0.1	18.7	7.02	1.89	4452	1.39	4.91
15	Maskinonge River	Wetland	26.7	6.4	23.2	0.9	9.2	7.11	1.89	4522	0.98	3.68
17	Zephyr Creek	Wetland	5.3	1.8	68.4	0	10.1	7.06	1.74	4631	1.42	9.2
29	Emily Creek	Wetland	8.2	7.5	37.6	0.6	32.8	7.09	2.04	4183	0.93	3.07

composition varied between our three land cover categories (Fig. 1). Wetland streams were characterized by DOM with relatively higher humic composition (HIX) and aromaticity (SUVA<sub>254</sub>). Forested streams were characterized by a relatively more humic-like DOM, while agricultural streams were characterized by DOM of more recent microbial origin and elevated nutrient concentrations (Fig. 1). We also found a wide range in humic-like (C1, C2, C3), microbial like (C5), anthropogenic microbial humic-like (C6), protein-like (C7), and soil-fulvic acid-like (C4) DOM compounds based on our seven component PARAFAC model (Supporting Information Table S1, Fig. S2). In general, C3

was the most abundant and most variable DOM component across the sites, whereas C7 had the lowest relative abundance. C6 ranged between 10% and 23% across streams but was more predominant in streams draining agricultural catchments. C1 and C2 were both highly abundant (20–26% and 15–26%, respectively) and were relatively invariant across the streams.

### Taxonomic composition

*Proteobacteria* and *Bacteroidetes* were most abundant across all sites (Supporting Information Fig. S3). We found no significant differences in taxonomic composition between sites



**Fig. 3.** (a) Ternary plots visualizing functions ( $n = 5526$ ) across the three land cover categories. Colors indicate gene modules as identified by WGCNA (Langfelder and Horvath 2008). Functions of which more than 60% were detected per land cover category were classified as being specialized to a certain land cover or DOM composition ( $n = 549$ ; indicated by the black lines). Circle size indicates relative abundance of the function and gray circles represent functions contained in modules that were not significantly related to land cover. Significant correlations (as derived from the WGCNA analysis) are indicated within the respective module; negative correlations are depicted by square brackets. Relative abundance of (b) gene modules (relationships with environmental variables are indicated in the boxes and correspond to panel a) and (c) metabolism types that are specialized to the respective land cover category.

(ANOVA, Bray–Curtis dissimilarity,  $r^2 = 0.1$ ,  $p > 0.05$ ). However, ternary plots revealed that the number of specialist microbial families decreased from natural to agricultural catchments (Supporting Information Fig. S3). Streams in forested catchments had the largest number of specialists ( $n = 7$ ), containing more specialists compared to agricultural and wetland streams ( $n = 4$  and  $n = 3$ , respectively). Families specialized to forested catchments include *Bacillaceae* and *Caulobacteraceae*. On the other hand, specialists found within wetland catchments include, *Geobacteraceae*, *Rhodocyclaceae*, *Lachnospiraceae*, and *Nitrosomonadaceae*. In contrast, agricultural streams were dominated by *Beijerinckiaceae* and *Sphingobacteriaceae*.

### Linking functional composition to DOM/nutrient composition

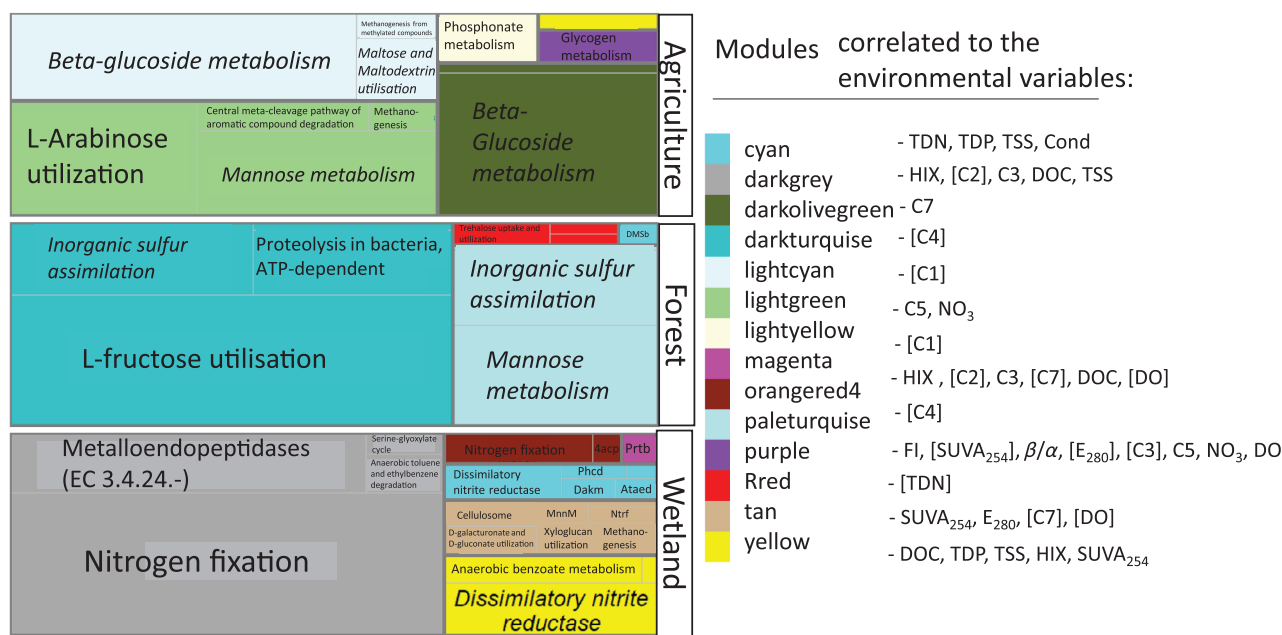
Functions related to protein, carbohydrate, and amino acid metabolism were most abundant across all streams (Supporting Information Fig. S5). Microbial functions differed significantly between stream sites (ANOVA using distance matrices, Bray–Curtis dissimilarity,  $r^2 = 0.3$ ,  $p < 0.05$ , Fig. 2a). Functional diversity was highest in forested streams and the number of unique functions was highest in forested and wetland streams ranging from 0.6% to 1.7% (Fig. 2b,c, Table 1).

WGCNA identified 53 modules containing anywhere from 20 to 816 functional genes. From these modules, 20 were related to DOM composition and/or nutrient concentrations ( $r > 0.62$ ,  $p < 0.05$ ; Supporting Information Fig. S6). Each of these modules was associated with 20–175 functional genes. For each gene in a module, module membership and GS were used to identify

which genes are important within the module and have the most relevance to nutrients and DOM composition.

We were able to identify a core set of functions that were not related to any of our agriculture, forested, or wetland categories, but rather were found in all three land cover types. Examples include functions related to DNA metabolism (e.g., DNA repair), protein folding, and biosynthesis (e.g., protein chaperones, universal GTPases), as well as biosynthesis of amino acids and derivatives (arginine biosynthesis, histidine biosynthesis) and central carbohydrate metabolism (e.g., glyoxylate bypass, pentose phosphate pathway, and glycolysis and gluconeogenesis). Although most microbial traits were found in all streams, we found several less abundant traits that were specialized to the individual ecosystems (Fig. 3a).

We identified clusters of traits related to specific DOM composition, directly driven by the respective land cover category. For instance, we observed higher abundances of phosphorus metabolism and membrane transport genes in agricultural streams which were characterized by microbially processed, protein-like DOM and higher nutrient concentrations. Furthermore, gene modules in agricultural streams were related to the synthesis and processing of DOM and proteins. Agricultural sites were further characterized by higher relative contribution of modules related to  $\text{NO}_3^-$ , TDP, and protein-like, freshly produced DOM (Fig. 3b). Within these modules, functions responsible for monomer uptake (e.g., mannose metabolism, l-arabinose utilization) and functions related to phosphorus cycling (e.g., phosphonate metabolism) were dominant (Fig. 4).



**Fig. 4.** Contribution of modules that are related to C, N, and P cycling and are specialized to the respective land cover category (374 individual functions). Significant correlations (as derived from the WGCNA analysis) are indicated within the respective module; square brackets indicate negative correlations. Module size indicates relative abundance within the respective land cover category, the color corresponds to the color derived from the WGCNA analysis. Modules and functions less abundant than 0.5% are abbreviated and listed in Supporting Information Table S4.



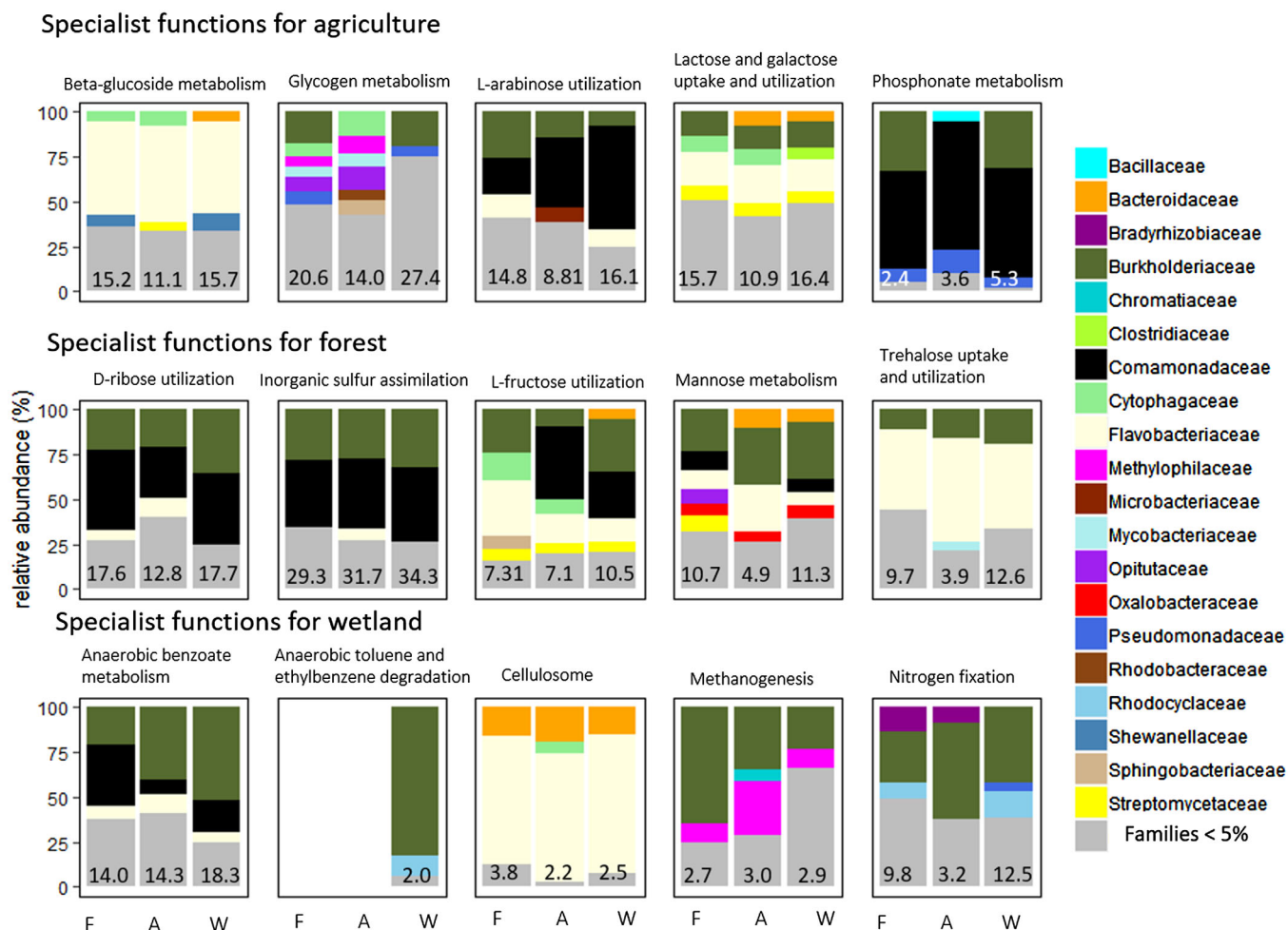
Streams draining wetlands were characterized by functions related to nitrogen metabolism and metabolism related to aromatic compounds, while forested streams were characterized by functions related to monomer uptake and photosynthesis (Fig. 3c). Streams with a greater abundance of humic-like and aromatic DOM, as is typical for wetlands, had a greater abundance of gene functions related to the metabolism of aromatic compounds (e.g., anaerobic toluene and ethylbenzene degradation, anaerobic benzoate metabolism, and anaerobic toluene metabolism). For instance, genes active in the processing and synthesis of aromatic/organic compounds (as indicated by HIX, SUVA<sub>254</sub>, and  $\epsilon$ 280) were more abundant in wetland streams. Functional modules specific to wetland streams were related to SUVA<sub>254</sub>, HIX, terrestrial humic-like fluorescence, TSS, DOC, TDN, and TDP ( $p < 0.05$ ).

In forested streams, only 29% of the genes most abundant in these catchments (above the 60% threshold) were significantly related to our measured environmental variables, while this number increased to 55% in agricultural and wetland catchments. The indicator species analysis identified indicator

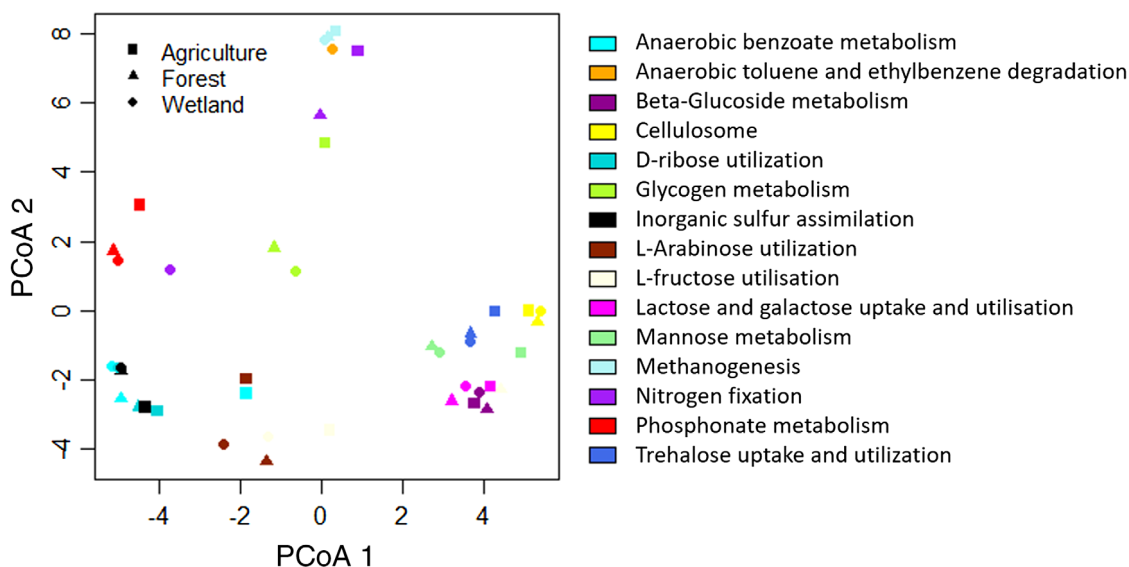
functions for each of the land cover categories (Supporting Information Fig. S7). These functions were largely coherent with the ones identified by the ternary plots. For instance, functions related to the metabolism of monosaccharides were abundant in agricultural streams, while nitrogen fixation was identified as an indicator function for wetland streams.

### Functional redundancy

We found that the most abundant functions specialized to either forested, agriculture, or wetland streams were present in a diverse group of microorganisms (Fig. 5). For instance, the potential for nitrogen fixation in wetlands is present in many relatively low abundance groups (< 5%) and few dominant ones, while in agricultural streams this function is dominated by two relatively abundant families, *Burkholderiaceae* and *Bradyrhizobiaceae*. As a second example, glycogen metabolism, a function related to energy storage, is dominated by *Cytophagaceae* and *Optitauaceae* in agricultural streams, and, *Burkholderiaceae* and *Pseudomonadaceae* in forested and wetland streams with a large contribution of microbial groups of low relative abundance. (Supporting Information Table S3).



**Fig. 5.** The taxonomic composition per selected “specialist function” ( $n = 15$ ) per land cover category (F: forest, A: agriculture, W: wetland). The inverted Simpson index is indicated in white for the families with a relative abundance below 5% (gray bars).



**Fig. 6.** PCoA based on a Bray–Curtis dissimilarity of the community composition per selected “specialist function” ( $n = 15$ ) per land cover category (a PCoA of all functions can be found in the Supporting Information) ( $n = 3$ ).

Functions related to the anaerobic degradation of toluene and ethylbenzene degradation) were only found in wetland-dominated streams. Other functions (e.g., phosphonate metabolism and trehalose uptake and utilization) were relatively conserved in their taxonomic composition in forested and wetland streams. There was no clear pattern in alpha diversity (as determined using the inversed Simpson index) between the taxonomic composition of generalist and specialist functions (Supporting Information Fig. S8). The taxonomic composition that was linked to a certain function varied in diversity within and across all land cover categories (Fig. 5). However, taxonomic richness per function was lowest in agricultural streams, compared to forested and wetland dominated streams where the richness was on average 1.38 and twofold higher per function, respectively (paired  $t$ -test,  $n = 168$ , corrected  $p$  value  $< 0.001$ ). PCoA shows that microbial community composition linked to the same function often differed substantially between land cover categories, that is, different taxa perform the same function in different environments (Figs. 5, 6 and Supporting Information Fig. S9). In contrast, different functions sometimes exhibited a higher taxonomic similarity within the same land cover category, that is, the same organisms perform multiple functions. (Figs. 5, 6 and Supporting Information Fig. S9). For instance, the taxonomic composition that was linked to anaerobic benzoate metabolism differed between the three land cover categories, and instead was similar to the taxonomic compositions linked to other functions within agricultural streams (e.g., L-arabinose utilization). On the other hand, a similar taxonomic community was linked to lactose and galactose uptake and utilization in wetland and forested streams. Thus, for some functions, the similarity of the microbial taxonomic compositions of different functions within a land cover category was higher than within the same function (e.g., nitrogen metabolism).

## Discussion

Our study uncovers patterns of stream microbial functional and taxonomic diversity in catchments ranging from natural (forest, wetland) to human-disturbed (croplands). We found that the number of specialist microbial families increased from agriculture to natural catchments. We also found that functional diversity and the number of uniquely occurring functions is higher in natural catchments, likely driven by specialized functions that are needed to capture the range in DOM composition and nutrients found in the wetland and forested streams.

DOM composition ranged from relatively humic-like, more complex DOM (C3) of high aromaticity in forested and wetland streams to more anthropogenic microbial humic-like DOM contributions in agricultural streams (C6). This may explain the comparatively lower functional diversity, but nonetheless distinct functional composition driven by elevated inputs of nutrients and freshly produced, more microbial-like DOM. This freshly produced and more microbial-like DOM is considered to be chemically less complex in quality compared to aromatic DOM which may constrain microbial functional diversity, leading to limited microbial resource niche differentiation (Hunting et al. 2017). In contrast, in natural (forest and wetland) catchments, it is likely that higher concentrations of complex humic-like DOM allow for a greater extent of niche differentiation and specialization and thus higher microbial functional diversity (Wilhelm et al. 2015). Altogether, our findings are consistent with the ecological theory that proposes that resource diversity can impact specialization and may increase species richness (Kassen 2002; Evans et al. 2005).

We found distinct functions that differed among agriculture, forested, and wetland-rich catchments, but nonetheless a considerable number of microbial traits were also conserved across the

landscape. In fact, genes involved in some of the core pathways were similarly abundant across all sampling sites. For example, we found that energy obtaining metabolic pathways of universal functions related to aerobic respiration and central carbohydrate metabolism (e.g., ATP-dependent Clp protease, ATP-binding subunit, glycolysis, and gluconeogenesis) were prevalent across all land cover categories. This core set of microbial functions is linked to biogeochemical cycles of carbon, nitrogen, oxygen, and sulfur and thus determines the structure of many microbial ecosystems (Falkowski et al. 2008). This similarity in functional gene composition across sites with differing environmental features may suggest the existence of functional redundancy to some extent. Thus, in the likely event an ecosystem becomes disturbed, there may be a potential for its key ecological processes to be restored by the reservoir of available microbes. However, the presence of functional genes does not necessarily mean that these genes are actively expressed. To confirm whether this is the case, further research is needed, which likely involves the simultaneous collection of data on gene occurrence, gene expression, and enzymatic activity.

#### Linking functional composition to DOM/nutrient composition

We identified specific links between DOM composition, nutrient concentrations, and the functional composition of the microbial community, highlighting the role of environmental forcing in shaping microbial functional traits. Nitrogen and phosphorus concentrations in agriculture- or wetland-dominated streams were strongly correlated to the microbial functional composition. Streams with high N and P concentrations also had higher relative abundances of genes involved in nutrient processing metabolic pathways such as the phosphonate metabolism and denitrification. Similarly, nutrient processing functions are well documented in agricultural streams (Opdyke et al. 2006; Ulén et al. 2007). Related to this, in catchments with high proportions of agriculture, we found gene modules that were positively correlated to microbially processed DOM (components C5 and C6). Microbially processed DOM has previously been shown to be related to human-dominated environments (Williams et al. 2016) and was found in high contributions in our agricultural streams (Wilson and Xenopoulos 2009; Williams et al. 2010).

Our study is an attempt to investigate the involvement and role of functional genes in carbon cycling and DOM composition in streams across land cover categories. We found linkages between DOM composition and functional genes. Proteins that were involved in the synthesis of amino acids were correlated with protein-like DOM (C7), an indication that they may play a role in its synthesis. Protein-like DOM (C7) fluorophores have been previously shown to increase microbial activity in the same study streams (Williams et al. 2010), and were hypothesized to be produced from extracellular peptide hydrolysis, or to provide a substrate that promotes extracellular aminopeptidase enzyme synthesis and activity (Williams et al. 2010). Additionally, genes involved in the processing of amino acids, specifically

tryptophan halogenase *PmA*, were observed in all our study streams but were particularly abundant in the agriculture streams. The tryptophan halogenase enzyme is essential in the biosynthesis of pyrrolnitrin (an antifungal antibiotic) by catalyzing the regio-selective chlorination of the 7-position of tryptophan, an amino acid that has the spectral signature of PARAFAC compound C7 (Dong et al. 2005; Williams et al. 2013; Karabencheva-Christova et al. 2017). Due to protein-like DOM being rich in nitrogen, microorganisms prefer to utilize these compounds to meet their energy and nutrient demands (Williams et al. 2010; Fuß et al. 2017). In fact, there may be a high turnover rate of these protein-like compounds, which have been linked to a higher abundance of bioavailable DOC (Guillemette and del Giorgio 2012). In contrast, C1, C2, and C3, which are more humic in nature, are more resistant to decomposition due to the extensive presence of condensed aromatic moieties and the high C:N ratios. Thus, they are a less preferential source of carbon for microbes than protein-rich compounds (McKnight et al. 2001; Guillemette and del Giorgio 2011). These humic-like components (e.g., C3) were instead associated with functions related to the metabolism of aromatic compounds (e.g., anaerobic toluene and ethylbenzene degradation).

#### Microbial traits differ between landscapes

The functional composition of the stream microbes differed distinctly between agriculture, forested, and wetland streams (Fig. 2). For instance, genes involved in the synthesis and processing of aromatic compounds (e.g., anaerobic toluene and ethylbenzene degradation, anaerobic benzoate metabolism) were prevalent in wetland streams which were characterized by humic-like and aromatic DOM (Supporting Information - Table S4). Wetland-dominated streams had high relative abundances of functions related to fermentation which are likely a result of anoxic/hypoxic conditions in parts of the wetland streams (Supporting Information Fig. S7). Additionally, in the presence of low oxygen conditions, we found that functions involved in nitrogen fixation (e.g., nitrogenase stabilizing/protective protein NifW, nitrogenase cofactor carrier protein NafY, probable iron binding protein from the HesB\_IscA\_SufA family in Nif operon), a process often associated with wetlands (Roger and Ladha 1992) were abundant in wetland streams. Other functions which we found to dominate in wetland streams include those related to potassium metabolism which may indicate microbial adaptation to high salinity as we observed in the wetland-dominated streams. Potassium metabolism was also identified as an indicator function for wetland streams using the LefSE method (Supporting Information Fig. S7).

Despite the extensive diversity in functional genes in forested catchments, 76% of the most abundant genes were not significantly related to any of our measured environmental variables. This may be because specialized functions were simply not related to any limnological variables that we measured. Two interesting findings are noticeable in the forest streams. First, we found elevated number of genes involved in

monomer uptake, sugar alcohols, and carbohydrate metabolism in our forested streams. These streams were characterized by low DOC concentrations, but high proportion of humic and refractory DOM and we postulate that microbial communities likely had more copies of carbohydrate processing genes to utilize the dominant DOM form efficiently. Second, we found that photosynthesis was a major metabolic function in forested and also wetland-dominated streams. There could be several reasons for this increase in photosynthetic genes related to either low or high light conditions both of which are possible in forested streams. Shading by trees would cause lower light availability but the relatively nonturbid nature of forested stream water (compared to our agriculture streams from long-term data; M. A. Xenopoulos unpubl.) would increase light availability. Both conditions could increase copies of photosynthetic genes required for either light utilization efficiency (under low light) or to take advantage of the high light conditions to increase photosynthetic rates. We further found functions related to sulfur metabolism to dominate in our forested streams, specifically the assimilation of inorganic sulfur, which may be a result of the low availability of organic matter (Fig. 1) and the resulting need to assimilate inorganic sulfur. Since sulfate is relatively rare in freshwaters, the enrichment in genes for the assimilation of inorganic sulfur may point to infiltration of sulfate-rich groundwaters to the forest streams.

We found that genes that correlated with microbially processed humic-like DOM (C6), and with P and N were less abundant in natural catchments as opposed to agricultural streams. This is not surprising as agricultural streams are richer in nutrients, have structurally less complex DOM and higher bacterial production (Wilson and Xenopoulos 2009; Bernot et al. 2010; Williams et al. 2010; Fuß et al. 2017). In line with our findings, high nutrient environments typically translate to high numbers of nutrient processing genes. For example, Medeiros et al. (2014) found the molecular markers *napA*, *amoA*, and *nfrA* accumulated in nitrogen-rich urban streams likely caused by the higher content of nutrients metabolized by these enzymes. Similarly, increased inputs of nutrients may explain the higher abundance of functions related to phosphorus metabolism in our agricultural streams. Also abundant in our agriculture streams was the DOM component C6 which was previously found to be highly correlated with population density and anthropogenic catchments (Williams et al. 2016). Interestingly in our study, C6 was correlated with a large number of functions related to urban activity. These functions include the vibrio core oligosaccharide biosynthesis and Heme, heme uptake, and utilization systems in gram-positives. This may also indicate the potential import of microbes from environments with urban activity.

In agricultural streams, we found high abundance of functions related to monomer uptake which may reflect the availability of relatively labile and fresh DOM or indicate the absence of specialized functions. A higher proportion of generalist functions like

nutrient and monomer uptake may result from the large fluctuation in nutrient and temperature regimes, which tend to support the dominance of generalists (Townsend 1989; Kassen 2002). These fluctuations can occur in agricultural streams (Petry et al. 2002; Fasching et al. 2018), particularly when fertilizers are applied or during storm events. Furthermore, our agricultural streams have flashier discharge hydrographs, thus potentially increasing stress on the microbes. Accordingly, genes involved in stress response and glycogen metabolism were more abundant in agricultural streams possibly indicating a microbial response to the environmental pressures (e.g., as an energy reserve to cope with environmental fluctuations). The fact that we predominantly observed generalists traits in agricultural streams may indicate that these traits are an adaptive response of stream microbial communities to maintain productivity and stability in these highly disturbed streams (Evans et al. 2005; Matias et al. 2013).

### Taxonomic composition

Overall, we found that forested streams had a greater number of specialists, both in taxonomy and gene functionality in comparison to agricultural catchments which hold a greater number of generalists indicating that there is a loss of biodiversity in streams with increasing anthropogenic activity. The families *Cryomorphaceae*, *Comamonadaceae*, *Flavobacteriaceae*, and *Legionellaceae* were the most abundant across all sample sites. These families encompass a wide range of functionally diverse microbes that are found in various aquatic ecosystems relatively rich in organic carbon due to their ability to maintain community productivity and stability (Bowman 2014; Lory 2014; McBride 2014; Willems 2014). Many of these families, such as *Comamonadaceae*, have previously been identified as generalists within stream communities (Wilhelm et al. 2015). Generalists are better equipped to cope with noticeable changes in the environment, such as changes in land use for example and resource availability (Kassen 2002; Wilhelm et al. 2015). This universal ability to cope with fluctuations in the environment seems to be caused by mechanisms of rapid functional adaptation such as horizontal gene transfers. Furthermore, such mechanisms may result in the rapid evolution of microorganisms and may explain the presence of specific species, even after streams have been subjected to changes following urbanization for example.

Noteworthy families potentially specialized to our forested catchments include *Lachnospiraceae*, *Peptococcaceae*, and *Bacillaceae*. The *Lachnospiraceae* family consists of anaerobic, fermentative, and chemoorganotrophic members with strong capability of hydrolyzing various organic compounds such as pectin,  $\alpha$ - and  $\beta$ -galactose,  $\alpha$ -amylose, and mannose (Stackebrandt 2014a; Haas and Blanchard 2017). This appears to further reaffirm our results (Fig. 5) that mannose metabolism is dominant in forested catchments and has been shown to be present in the soils of forested catchments. Furthermore, *Peptococcaceae* is a phylogenetically heterogeneous family which contains many members involved in the assimilation and subsequent reduction of inorganic sulfur (e.g., sulfate, sulfite, and

thiosulfate) (Stackebrandt 2014b). In line with this, functions related to sulfur metabolism were elevated in our forested streams which may be also caused by the abundance of *Peptococcaceae*. Last, *Bacillaceae* has been shown to be a functionally diverse family, detected in a broad range of habitats from freshwaters to oceans, involved in the breakdown of cellulose, hemicellulose, and pectin, suggesting that they contribute in the degradation and mineralization of plant and humic materials and overall biogeochemical cycling of carbon (Maki et al. 2012; Mandic-Mulec et al. 2016).

In our study, wetland-dominated streams included several microbial specialists such as *Nitrosomonadaceae*, *Rhodocyclaceae*, and *Geobacteraceae*. These streams were also characterized by a higher proportion of the humic-like C3 DOM component which may explain the presence of *Rhodocyclaceae* known to degrade aromatic organic compounds. Within the *Rhodocyclaceae*, the genus *Azospira* is able to use humic acids as electron donors in agricultural soils (Van Trump et al. 2011). Additionally, the *Geobacteraceae* family has been shown to utilize small organic acids and alcohols with several species capable of oxidizing aromatic hydrocarbons, such as toluene and benzene (Röling 2014). Also abundant in our wetland sites, *Nitrosomonadaceae*—a phylogenetic group that has widely been associated with increased anthropogenic nitrogen loading (Medeiros et al. 2014). In our wetland catchments, we found slightly elevated TDN concentrations, which may explain the higher abundance of *Nitrosomonadaceae*. Many members are lithoautotrophic ammonia oxidizers converting ammonia to nitrite, which is then subsequently oxidized to nitrate by other microbes. Thus, this family is essential to nitrification/denitrification and the nitrogen cycle as a whole in freshwater ecosystems (Prosser et al. 2014).

We found lower abundance of specialist microbial families in agricultural streams relative to forested catchments which may be linked to the stream channel homogeneity and flashier discharge, as previously discussed. In agricultural catchments, many microbial families may be washed into the streams from soils during high flows when agricultural soils are susceptible to erosion (Meyer and Harmon 1984) which may explain why we find a greater number of soil-related taxa. Agricultural streams were dominated by families such as *Acidimicrobiaceae* which previously were reported to play a key role in the oxidation of ammonium or iron (Stackebrandt 2014c; Huang and Jaffé 2018). Although the *Acidimicrobiaceae* family is generally associated with wetland catchments due to their preference for acidic environments, functions related to the acquisition and metabolism of iron were predominant within our agricultural streams and may be linked to the occurrence of this family (Fig. 3c). The family *Beijerinckiaceae*, abundant in acidic soils, consists of members which are chemo-organotrophic or methanotrophic and aerobic, but may also utilize a broad spectrum of organic compounds (Marín and Arahál 2014). Another abundant family in our agriculture streams, *Sphingobacteriaceae*, have also been widely observed in soils and compost and are believed to be deeply involved in carbon cycling and the degradation and conversion of organic molecules (Lambiase 2014).

### Functional redundancy

Our results indicate that functional redundancy may be important for stream ecosystems in a mixed land use setting. The degree of functional redundancy may vary based on the environmental conditions and the functions considered (Louca et al. 2018). We found evidence for functional redundancy across all land cover categories as several microbial taxa were linked to the same function, while species richness per function was highest in natural catchments. This may indicate that functions of natural catchments are potentially more resilient to disturbance and species loss. Additionally, microbial community composition linked to specific functions often differed substantially between land cover categories indicating that there is potential for functional similarity (Allison and Martiny 2008) between communities from different land cover categories. However, while some functions were dominated by different microbial families (e.g., anaerobic benzoate metabolism), others remained relatively conserved in their community composition (e.g., phosphonate metabolism, trehalose uptake, and utilization). These contrasting findings are in line with previous studies showing functional redundancy appears to be dependent on various environmental factors (Comte et al. 2013b; Peter and Sommaruga 2016), while metabolic plasticity can be an intrinsic property of bacterial communities (Comte et al. 2013b). In contrast, other studies found a lack of functional redundancy in freshwater systems (Delgado-Baquerizo et al. 2016). Lending support to this finding, Peter et al. (2011) propose that species identity and community composition rather than richness is more important for microbial processes. Emerging patterns of functional redundancy may result from biotic interactions as well as environmental and spatial processes (Louca et al. 2018). This notion is supported by our results showing different degrees of functional redundancy. Microbial functional redundancy seemed to be prevalent for functions related to energy metabolism, like the uptake of simple sugars, which was an abundant function in forested and agricultural streams. Indeed, metabolic pathways involved in energy metabolism can be strongly coupled with certain environmental factors (Raes et al. 2011), and may be decoupled from the taxonomical composition (Louca et al. 2017). Since our results show the loss of functional diversity in agriculture streams, it appears that functional diversity to ensure important ecosystem processes can only be kept to an extent.

### Conclusion

Our results highlight the impact of the environment on microbial community structure and function. Although we show evidence for distinct changes in microbial functionality and the loss of functional diversity when comparing natural to agricultural streams, there is a certain degree of functional similarity across all streams which may allow communities to maintain key functions for the biogeochemical cycling of C, N, and P. However, this may not be the case for all functions, as indicated by the loss of functional diversity in agricultural

catchments. The high taxonomical richness per function in natural catchments may indicate that these catchments are potentially more resilient to disturbance and species loss. Species richness may be important for ecosystem functioning as it provides a pool of species with potentially relevant traits (Fetzer et al. 2015).

Our findings provide detailed insights into the relationship between microbial functions and biogeochemical processes in streams and how these biogeochemical processes are reflected by specific microbial communities. This knowledge forms a basis to better evaluate biogeochemical changes resulting from anthropogenic activities. This evaluation is needed for sustainable land management, especially to define thresholds at which ecosystems maintain resistance to environmental stress and do not move to another unwanted state with yet unknown ecological and biogeochemical implications.

## References

- Allison, S. D., and J. B. H. Martiny. 2008. Resistance, resilience, and redundancy in microbial communities. *Proc. Natl. Acad. Sci. USA* **105**: 11512–11519.
- American Public Health Association. 1992. Standard methods for the examination of water and wastewater. 18th edn. Washington D.C: American Public Health Association.
- Andersson, C. A., and R. Bro. 2000. The N-way toolbox for MATLAB. *Chemometr. Intell. Lab. Syst.* **52**: 1–4. doi:10.1016/S0169-7439(00)00071-X
- Bankevich, A., and others. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**: 455–477. doi:10.1089/cmb.2012.0021
- Battin, T. J., L. A. Kaplan, J. Denis Newbold, and C. M. E. Hansen. 2003. Contributions of microbial biofilms to ecosystem processes in stream mesocosms. *Nature* **426**: 439–442. doi:10.1038/nature02152
- Battin, T. J., L. A. Kaplan, S. Findlay, C. S. Hopkinson, E. Marti, A. I. Packman, J. D. Newbold, and F. Sabater. 2008. Biophysical controls on organic carbon fluxes in fluvial networks. *Nat. Geosci.* **1**: 95–100. doi:10.1038/ngeo101
- Bernot, M. J., and others. 2010. Inter-regional comparison of land-use effects on stream metabolism. *Freshw. Biol.* **55**: 1874–1890. doi:10.1111/j.1365-2427.2010.02422.x
- Bowman, J. P. 2014. The family *Cryomorpaceae*, p. 539–550. In Rosenberg, E., E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson [eds.], *The prokaryotes*. Springer.
- Bushnell, B. 2014. BBLMap: A fast, accurate, splice-aware aligner (No. LBNL-7065E). Lawrence Berkeley National Lab (LBNL).
- Chen, I.-M. A., and others. 2016. IMG/M: Integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res.* **45**: D507–D516. doi:10.1093/nar/gkw929
- Comte, J., L. Fauteux, and P. A. del Giorgio. 2013a. Links between metabolic plasticity and functional redundancy in freshwater bacterioplankton communities. *Front. Microbiol.* **4**: 1–11. doi:10.3389/fmicb.2013.00112
- Comte, J., P. A. Giorgio, and S. URL. 2013b. Linking the patterns of change in composition and function in bacterioplankton successions along environmental gradients. *Ecology* **91**: 1466–1476. doi:10.1890/09-0848.1
- D'Amario, S. C., and M. A. Xenopoulos. 2015. Linking dissolved carbon dioxide to dissolved organic matter quality in streams. *Biogeochemistry* **126**: 99–114. doi:10.1007/s10533-015-0143-y
- Delgado-Baquerizo, M., L. Giaramida, P. B. Reich, A. N. Khachane, K. Hamonts, C. Edwards, L. A. Lawton, and B. K. Singh. 2016. Lack of functional redundancy in the relationship between microbial diversity and ecosystem functioning. *J. Ecol.* **104**: 936–946. doi:10.1111/1365-2745.12585
- Delmont, T. O., C. Quince, A. Shaiber, O. C. Esen, S. T. M. Lee, S. Lucker, and A. M. Eren. 2018. Nitrogen-fixing populations of *Planctomycetes* and *Proteobacteria* are abundant in the surface ocean metagenomes. *Nat. Microbiol.*, **3**(7), 804.
- Dong, C., S. Flecks, S. Unversucht, C. Haupt, K.-H. van Pee, and J. H. Naismith. 2005. Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination. *Science* **309**: 2216–2219. doi:10.1126/science.1116510
- Dopheide, A., G. Lear, Z. He, J. Zhou, and G. D. Lewis. 2015. Functional gene composition, diversity and redundancy in microbial stream biofilm communities. *PLoS One* **10**: 1–21. doi:10.1371/journal.pone.0123179
- Evans, K. L., P. H. Warren, and K. J. Gaston. 2005. Species–energy relationships at the macroecological scale: A review of the mechanisms. *Biol. Rev.* **80**: 1–25.
- Falkowski, P. G., T. Fenchel, and E. F. Delong. 2008. The microbial engines that drive Earth's biogeochemical cycles. *Science* **320**: 1034–1039. doi:10.1126/science.1153213
- Fasching, C., H. F. Wilson, S. C. D'Amario, and M. A. Xenopoulos. 2018. Natural land cover in agricultural catchments alters flood effects on DOM composition and decreases nutrient levels in streams. *Ecosystems*: 1–16. doi:10.1007/s10021-019-00354-0
- Fetzer, I., K. Johst, R. Schäwe, T. Banitz, H. Harms, and A. Chatzinotas. 2015. The extent of functional redundancy changes as species' roles shift in different environments. *Proc. Natl. Acad. Sci. USA* **112**: 14888–14893. doi:10.1073/pnas.1505587112
- Findlay, S. 2010. Stream microbial ecology. *J. North Am. Benthol. Soc.* **29**: 170–181. doi:10.1899/09-023.1
- Fuß, T., B. Behounek, A. J. Ulseth, and G. A. Singer. 2017. Land use controls stream ecosystem metabolism by shifting dissolved organic matter and nutrient regimes. *Freshw. Biol.* **62**: 582–599. doi:10.1111/fwb.12887
- Green, J. L., B. J. M. Bohannan, and R. J. Whitaker. 2008. Microbial biogeography: From taxonomy to traits. *Science* **320**: 1039–1043. doi:10.1126/science.1153475

- Gtari, M., F. Ghodhbane-Gtari, I. Nouioui, N. Beauchemin, and L. S. Tisa. 2012. Phylogenetic perspectives of nitrogen-fixing actinobacteria. *Arch. Microbiol.* **194**: 3–11. doi:[10.1007/s00203-011-0733-6](https://doi.org/10.1007/s00203-011-0733-6)
- Guillemette, F., and P. A. del Giorgio. 2011. Reconstructing the various facets of dissolved organic carbon bioavailability in freshwater ecosystems. *Limnol. Oceanogr.* **56**: 734–748. doi:[10.4319/lo.2011.56.2.0734](https://doi.org/10.4319/lo.2011.56.2.0734)
- Guillemette, F., and P. A. del Giorgio. 2012. Simultaneous consumption and production of fluorescent dissolved organic matter by lake bacterioplankton. *Environ. Microbiol.* **14**: 1432–1443. doi:[10.1111/j.1462-2920.2012.02728.x](https://doi.org/10.1111/j.1462-2920.2012.02728.x)
- Haas, K. N., and J. L. Blanchard. 2017. *Kineothrix alysoides*, gen. nov., sp. nov., a saccharolytic butyrate-producer within the family *Lachnospiraceae*. *Int. J. Syst. Evol. Microbiol.* **67**: 402–410. doi:[10.1099/ijsem.0.001643](https://doi.org/10.1099/ijsem.0.001643)
- Hosen, J. D., C. M. Febria, B. C. Crump, and M. A. Palmer. 2017. Watershed urbanization linked to differences in stream bacterial community composition. *Front. Microbiol.* **8**: 1–17. doi:[10.3389/fmicb.2017.01452](https://doi.org/10.3389/fmicb.2017.01452)
- Huang, S., and P. R. Jaffé. 2018. Isolation and characterization of an ammonium-oxidizing iron reducer: *Acidimicrobiaceae* sp. A6. *PLoS One* **13**: e0194007. doi:[10.1371/journal.pone.0194007](https://doi.org/10.1371/journal.pone.0194007)
- Hunting, E. R., S. H. Barmantlo, M. Schrama, P. M. van Bodegom, Y. Zhai, and M. G. Vijver. 2017. Agricultural constraints on microbial resource use and niche breadth in drainage ditches. *PeerJ* **5**: e4175. doi:[10.7717/peerj.4175](https://doi.org/10.7717/peerj.4175)
- Karabancheva-Christova, T. G., J. Torras, A. J. Mulholland, A. Lodola, and C. Z. Christov. 2017. Mechanistic insights into the reaction of chlorination of tryptophan catalyzed by tryptophan 7-halogenase. *Sci. Rep.* **7**: 17395. doi:[10.1038/s41598-017-17789-x](https://doi.org/10.1038/s41598-017-17789-x)
- Kassen, R. 2002. The experimental evolution of specialists, generalists, and the maintenance of diversity. *J. Evol. Biol.* **15**: 173–190. doi:[10.1046/j.1420-9101.2002.00377.x](https://doi.org/10.1046/j.1420-9101.2002.00377.x)
- Keegan, K. P., E. M. Glass, and F. Meyer. 2016. MG-RAST, a metagenomics service for analysis of microbial community structure and function, p. 207–233. *In* *Microbial environmental genomics (MEG)*. Humana Press.
- Lambiase, A. 2014. The family sphingobacteriaceae, p. 907–914. *In* Rosenberg E., E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson [eds.], *The prokaryotes*. Springer.
- Langenheder, S., E. S. Lindström, and L. J. Tranvik. 2006. Structure and function of bacterial communities emerging from different sources under identical conditions. *Appl. Environ. Microbiol.* **72**: 212–220. doi:[10.1128/AEM.72.1.212-220.2006](https://doi.org/10.1128/AEM.72.1.212-220.2006)
- Langfelder, P., and S. Horvath. 2008. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**: 559.
- LeBrun, E. S., R. S. King, J. A. Back, and S. Kang. 2018. Microbial community structure and function decoupling across a phosphorus gradient in streams. *Microb. Ecol.* **75**: 64–73. doi:[10.1007/s00248-017-1039-2](https://doi.org/10.1007/s00248-017-1039-2)
- Lory, S. 2014. The family Legionellaceae, p. 387–389. *In* Rosenberg E., E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson [eds.], *The prokaryotes*. Springer.
- Louca, S., S. M. S. Jacques, A. P. F. Pires, J. S. Leal, D. S. Srivastava, L. W. Parfrey, V. F. Farjalla, and M. Doebeli. 2017. High taxonomic variability despite stable functional structure across microbial communities. *Nat. Ecol. Evol.* **1**: 15.
- Louca, S., and others. 2018. Function and functional redundancy in microbial systems. *Nat. Ecol. Evol.* **2**(6), 936. doi:[10.1038/s41559-018-0519-1](https://doi.org/10.1038/s41559-018-0519-1)
- Maki, M. L., A. Idrees, K. T. Leung, and W. Qin. 2012. Newly isolated and characterized bacteria with great application potential for decomposition of lignocellulosic biomass. *J. Mol. Microbiol. Biotechnol.* **22**: 156–166. doi:[10.1159/000341107](https://doi.org/10.1159/000341107)
- Mandic-Mulec, I., P. Stefanic, and J. D. van Elsas. 2016. Ecology of bacillaceae, p. 59–85. *In* *The bacterial spore: From molecules to systems*. American Society of Microbiology.
- Marín, I., and D. R. Arahal. 2014. The family beijerinckiaceae, p. 115–133. *In* Rosenberg, E., E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson [eds.], *The prokaryotes*. Springer.
- Matias, M., M. Combe, C. Barbera, and N. Mouquet. 2013. Ecological strategies shape the insurance potential of biodiversity. *Front. Microbiol.* **3**: 432.
- McBride, M. J. 2014. The family flavobacteriaceae, p. 643–676. *In* Rosenberg, E., E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson [eds.], *The prokaryotes*. Springer.
- McKnight, D. M., E. W. Boyer, P. K. Westerhoff, P. T. Doran, T. Kulbe, and D. T. Anderson. 2001. Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. *Limnol. Oceanogr.* **46**: 38–48. doi:[10.4319/lo.2001.46.1.0038](https://doi.org/10.4319/lo.2001.46.1.0038)
- Medeiros, J. D., L. X. Araújo, V. L. da Silva, C. G. Diniz, D. E. Cesar, A. Del'Duca, and C. M. Coelho. 2014. Characterization of the microbial community in a lotic environment to assess the effect of pollution on nitrifying and potentially pathogenic bacteria. *Braz. J. Biol.* **74**: 612–622. doi:[10.1590/1519-6984.26712](https://doi.org/10.1590/1519-6984.26712)
- Meyer, L. D., and W. C. Harmon. 1984. Susceptibility of agricultural soils to interrill erosion. *Soil Sci. Soc. Am. J.* **48**: 1152–1157. doi:[10.2136/sssaj1984.03615995004800050040x](https://doi.org/10.2136/sssaj1984.03615995004800050040x)
- Newton, R. J., S. E. Jones, A. Eiler, K. D. McMahon, and S. Bertilsson. 2011. A guide to the natural history of freshwater lake bacteria. *Microbiol. Mol. Biol. Rev.* **75**: 14–49. doi:[10.1128/MMBR.00028-10](https://doi.org/10.1128/MMBR.00028-10)
- Nishimoto, K. P., D. Newkirk, S. Hou, J. Fruehauf, and E. L. Nelson. 2007. Fluorescence activated cell sorting (FACS) using RNAlater to minimize RNA degradation and perturbation of mRNA expression from cells involved in initial host-microbe interactions. *J. Microbiol. Methods* **70**: 205–208. doi:[10.1016/j.mimet.2007.03.022](https://doi.org/10.1016/j.mimet.2007.03.022)

- Oksanen, J., and others. 2013. Package 'vegan.' Community ecology package, version 2.
- Oksanen, J., and others. 2017. Vegan: Community ecology package. R package version 2.4-5.
- Opdyke, M. R., M. B. David, and B. L. Rhoads. 2006. Influence of geomorphological variability in channel characteristics on sediment denitrification in agricultural streams. *J. Environ. Qual.* **35**: 2103–2112. doi:10.2134/jeq2006.0072
- Osman, O. A., S. Beier, M. Grabherr, and S. Bertilsson. 2017. Interactions of freshwater cyanobacteria with bacterial antagonists. *Appl. Environ. Microbiol.* **83**: e02634–e02616.
- Osterholz, H., G. Singer, B. Wemheuer, R. Daniel, M. Simon, J. Niggemann, and T. Dittmar. 2016. Deciphering associations between dissolved organic molecules and bacterial communities in a pelagic marine system. *ISME J.* **10**: 1717–1730. doi:10.1038/ismej.2015.231
- Overbeek, R., and others. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* **33**: 5691–5702. doi:10.1093/nar/gki866
- Peter, H., S. Beier, S. Bertilsson, E. S. Lindström, S. Langenheder, and L. J. Tranvik. 2011. Function-specific response to depletion of microbial diversity. *ISME J.* **5**: 351–361. doi:10.1038/ismej.2010.119
- Peter, H., and R. Sommaruga. 2016. Shifts in diversity and function of lake bacterial communities upon glacier retreat. *ISME J.* **10**: 1545–1554. doi:10.1038/ismej.2015.245
- Petry, J., C. Soulsby, I. A. Malcolm, and A. F. Youngson. 2002. Hydrological controls on nutrient concentrations and fluxes in agricultural catchments. *Sci. Total Environ.* **294**: 95–110.
- Prosser, J. I., I. M. Head, and L. Y. Stein. 2014. The family nitrosomonadaceae, p. 901–918. *In* Rosenberg, E., E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson [eds.], *The prokaryotes*. Springer.
- R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing. [accessed 2018 August 15]. Available from <http://www.R-project.org/>
- Raes, J., I. Letunic, T. Yamada, L. J. Jensen, and P. Bork. 2011. Toward molecular trait-based ecology through integration of biogeochemical, geographical and metagenomic data. *Mol. Syst. Biol.* **7**: 473. doi:10.1038/msb.2011.6
- Roger, P. A., and Ladha, J. K. (1992). Biological N<sub>2</sub> fixation in wetland rice fields: Estimation and contribution to nitrogen balance. *In* *Biological Nitrogen Fixation for Sustainable Agriculture* (pp. 41–55). Springer.
- Röling, W. F. M. 2014. The family Geobacteraceae, p. 157–172. *In* Rosenberg, E., E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson [eds.], *The prokaryotes*. Springer.
- Ruiz-González, C., J. P. Niño-García, J.-F. Lapierre, and P. A. del Giorgio. 2015. The quality of organic matter shapes the functional biogeography of bacterioplankton across boreal freshwater ecosystems. *Glob. Ecol. Biogeogr.* **24**: 1487–1498. doi:10.1111/geb.12356
- Segata, N., J. Izard, L. Waldron, D. Gevers, L. Miropolsky, W. S. Garrett, and C. Huttenhower. 2011. Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**: R60. doi:10.1186/gb-2011-12-6-r60
- Sokol, R. R., and F. J. Rohlf. 1995. The principles and practice of statistics in biological research. p. 888. Freeman and Company.
- Stackebrandt, E. 2014a. The family Lachnospiraceae, p. 197–201. *In* Rosenberg, E., E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson [eds.], *The prokaryotes*. Springer.
- Stackebrandt, E. 2014b. The emended family Peptococcaceae and description of the families Desulfotobacteriaceae, Desulfotomaculaceae, and Thermincolaceae, p. 285–290. *In* Rosenberg, E., E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson [eds.], *The prokaryotes*. Springer.
- Stackebrandt, E. 2014c. The family Acidimicrobiaceae, p. 5–12. *In* Rosenberg, E., E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson [eds.], *The prokaryotes*. Springer.
- Stedmon, C. A., and R. Bro. 2008. Characterizing dissolved organic matter fluorescence with parallel factor analysis: A tutorial. *Limnol. Oceanogr.: Methods* **6**: 572–579. doi:10.4319/lom.2008.6.572
- Townsend, C. R. 1989. The patch dynamics concept of stream community ecology. *J. North Am. Benthol. Soc.* **8**: 36–50. doi:10.2307/1467400
- Traving, S. J., O. Rowe, N. M. Jakobsen, H. Sørensen, J. Dinasquet, C. A. Stedmon, A. Andersson, and L. Riemann. 2017. The effect of increased loads of dissolved organic matter on estuarine microbial community composition and function. *Front. Microbiol.* **8**: 351. doi:10.3389/fmicb.2017.00351
- Ulén, B., M. Bechmann, J. Fölster, H. P. Jarvie, and H. Tunney. 2007. Agriculture as a phosphorus source for eutrophication in the north-west European countries, Norway, Sweden, United Kingdom and Ireland: A review. *Soil Use Manag.* **23**: 5–15. doi:10.1111/j.1475-2743.2007.00115.x
- Vachon, D., Y. T. Prairie, F. Guillemette, and P. A. del Giorgio. 2017. Modeling allochthonous dissolved organic carbon mineralization under variable hydrologic regimes in boreal lakes. *Ecosystems* **20**: 781–795. doi:10.1007/s10021-016-0057-0
- Van Trump, J. I., K. C. Wrighton, J. C. Thrash, K. A. Weber, G. L. Andersen, and J. D. Coates. 2011. Humic acid-oxidizing, nitrate-reducing bacteria in agricultural soils. *MBio* **2**: e00044–e00011. doi:10.1128/mBio.00044-11
- Weishaar, J., G. Aiken, B. Bergamaschi, M. Fram, R. Fujii, and K. Mopper. 2003. Evaluation of specific ultra-violet absorbance as an indicator of the chemical content of dissolved organic carbon. *Environ. Sci. Technol.* **37**: 4702–4708. doi:10.1021/es030360x



- Wilhelm, L., G. A. Singer, C. Fasching, T. J. Battin, and K. Besemer. 2013. Microbial biodiversity in glacier-fed streams. *ISME J.* **7**: 1651–1660. doi:[10.1038/ismej.2013.44](https://doi.org/10.1038/ismej.2013.44)
- Wilhelm, L., K. Besemer, L. Fragner, H. Peter, W. Weckwerth, and T. J. Battin. 2015. Altitudinal patterns of diversity and functional traits of metabolically active microorganisms in stream biofilms. *ISME J.* **9**: 2454–2464. doi:[10.1038/ismej.2015.56](https://doi.org/10.1038/ismej.2015.56)
- Willems, A. 2014. The family Comamonadaceae, p. 777–851. *In* Rosenberg, E., E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson [eds.], *The prokaryotes*. Springer.
- Williams, C. J., Y. Yamashita, H. F. Wilson, R. Jaffe, and M. A. Xenopoulos. 2010. Unraveling the role of land use and microbial activity in shaping dissolved organic matter characteristics in stream ecosystems. *Limnol. Oceanogr.* **55**: 1159–1171. doi:[10.4319/lo.2010.55.3.1159](https://doi.org/10.4319/lo.2010.55.3.1159)
- Williams, C. J., P. C. Frost, and M. A. Xenopoulos. 2013. Beyond best management practices: Pelagic biogeochemical dynamics in urban stormwater ponds. *Ecol. Appl.* **23**: 1384–1395. doi:[10.1890/12-0825.1](https://doi.org/10.1890/12-0825.1)
- Williams, C. J., P. C. Frost, A. M. Morales-Williams, J. H. Larson, W. B. Richardson, A. S. Chianget, and M. A. Xenopoulos. 2016. Human activities cause distinct dissolved organic matter composition across freshwater ecosystems. *Glob. Chang. Biol.* **22**: 613–626. doi:[10.1111/gcb.13094](https://doi.org/10.1111/gcb.13094)
- Wilson, H. F., and M. A. Xenopoulos. 2008. Ecosystem and seasonal control of stream dissolved organic carbon along a gradient of land use. *Ecosystems* **11**: 555–568. doi:[10.1007/s10021-008-9142-3](https://doi.org/10.1007/s10021-008-9142-3)
- Wilson, H. F., and M. A. Xenopoulos. 2009. Effects of agricultural land use on the composition of fluvial dissolved organic matter. *Nat. Geosci.* **2**: 37–41. doi:[10.1038/ngeo391](https://doi.org/10.1038/ngeo391)
- Zeglin, L. H. 2015. Stream microbial diversity in response to environmental changes: Review and synthesis of existing research. *Front. Microbiol.* **6**: 454. doi:[10.3389/fmicb.2015.00454](https://doi.org/10.3389/fmicb.2015.00454)
- Zhao, M., K. Xue, F. Wang, S. Liu, S. Bai, B. Sun, J. Zhou, and Y. Yang. 2014. Microbial mediation of biogeochemical cycles revealed by simulation of global changes with soil transplant and cropping. *ISME J.* **8**: 2045–2055. doi:[10.1038/ismej.2014.46](https://doi.org/10.1038/ismej.2014.46)
- Zhou, J., M. A. Bruns, and J. M. Tiedje. 1996. DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **62**: 316–322.
- Zsolnay, A., E. Baigar, M. Jimenez, B. Steinweg, and F. Saccomandi. 1999. Differentiating with fluorescence spectroscopy the sources of dissolved organic matter in soils subjected to drying. *Chemosphere* **38**: 45–50. doi:[10.1016/S0045-6535\(98\)00166-0](https://doi.org/10.1016/S0045-6535(98)00166-0)

### Acknowledgments

We thank Andrew Scott who helped with water sampling and data extraction. We also thank Sarah King for her assistance in the PARAFAC analysis. This project was funded by Canada's Natural Sciences and Engineering Research Council (NSERC) Discovery Grant to M.A.X. and by the project BIBS provided by the German Ministry of Education and Science (BMBF; 01LC1501G) to H.-P.G. M.B. acknowledges financial support from Deutsche Forschungsgemeinschaft project (BI 1987/2-1).

### Conflict of Interest

None declared.

Submitted 31 January 2019

Revised 17 June 2019

Accepted 20 September 2019

Associate editor: Ramon Massana Molera