

## ORIGINAL ARTICLE

# Deciphering associations between dissolved organic molecules and bacterial communities in a pelagic marine system

Helena Osterholz<sup>1</sup>, Gabriel Singer<sup>2,3</sup>, Bernd Wemheuer<sup>4</sup>, Rolf Daniel<sup>4</sup>, Meinhard Simon<sup>1</sup>, Jutta Niggemann<sup>1</sup> and Thorsten Dittmar<sup>1</sup>

<sup>1</sup>Institute for Chemistry and Biology of the Marine Environment (ICBM), University of Oldenburg, Oldenburg, Germany; <sup>2</sup>Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin, Germany; <sup>3</sup>Hanse-Wissenschaftskolleg (HWK), Delmenhorst, Germany and <sup>4</sup>Institute of Microbiology and Genetics, Georg August University Göttingen, Göttingen, Germany

Dissolved organic matter (DOM) is the main substrate and energy source for heterotrophic bacterioplankton. To understand the interactions between DOM and the bacterial community (BC), it is important to identify the key factors on both sides in detail, chemically distinct moieties in DOM and the various bacterial taxa. Next-generation sequencing facilitates the classification of millions of reads of environmental DNA and RNA amplicons and ultrahigh-resolution mass spectrometry yields up to 10 000 DOM molecular formulae in a marine water sample. Linking this detailed biological and chemical information is a crucial first step toward a mechanistic understanding of the role of microorganisms in the marine carbon cycle. In this study, we interpreted the complex microbiological and molecular information via a novel combination of multivariate statistics. We were able to reveal distinct relationships between the key factors of organic matter cycling along a latitudinal transect across the North Sea. Total BC and DOM composition were mainly driven by mixing of distinct water masses and presumably retain their respective terrigenous imprint on similar timescales on their way through the North Sea. The active microbial community, however, was rather influenced by local events and correlated with specific DOM molecular formulae indicative of compounds that are easily degradable. These trends were most pronounced on the highest resolved level, that is, operationally defined 'species', reflecting the functional diversity of microorganisms at high taxonomic resolution. *The ISME Journal* (2016) 10, 1717–1730; doi:10.1038/ismej.2015.231; published online 22 January 2016

## Introduction

Marine bacterioplankton depend on the availability of dissolved organic matter (DOM) as their main substrate and energy source (Azam *et al.*, 1983; Fuhrman *et al.*, 2015). Interactions between heterotrophic organisms and DOM are manifold and shape DOM and microbial community composition. Microbially driven processes affecting DOM composition include selective uptake and consumption for growth, energy storage and metabolism, and excretion of secondary compounds and cleavage products (Dittmar and Stubbins, 2014). In order to understand the prevailing interactions between DOM and a bacterial community (hereafter BC), it is important to identify the key factors on both sides in detail.

There is growing evidence that the BC composition in the marine realm is driven to a large extent by deterministic selection through environmental factors (Stegen *et al.*, 2012; Valentin-Vargas *et al.*, 2012), among which substrate availability and composition of DOM rank prominently (Harvey *et al.*, 2006; Gomez-Consarnau *et al.*, 2012), but also physical factors have a role (Liu *et al.*, 2013). Closely related and functionally similar taxa are found across similar habitats more often than expected by chance (Kraft *et al.*, 2007; Teeling *et al.*, 2012) and substrate-specific allocation to the vast niche space offered by complex DOM may contribute to maintaining the highly diverse microbial communities in the oceans (Hutchinson, 1961; Zinger *et al.*, 2011; Gifford *et al.*, 2013).

Technical advances in molecular microbial ecology (Edwards and Dinsdale, 2007) and organic geochemistry (Kujawinski, 2002; Koch *et al.*, 2005) have enabled the ultrahigh-resolution analysis of both, the microbial community and DOM. Next-generation sequencing of marker genes facilitates the classification of millions of reads of

Correspondence: H Osterholz, Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, Carl-von-Ossietzky-Strasse 9-11, Oldenburg D-26129, Germany.  
E-mail: helena.osterholz@uni-oldenburg.de  
Received 4 June 2015; revised 26 October 2015; accepted 6 November 2015; published online 22 January 2016

environmental DNA and RNA amplicons and ultrahigh-resolution mass spectrometry (via the Fourier-transform ion cyclotron resonance technique, FT-ICR-MS) typically yields thousands of molecular formulae for one oceanic DOM sample. This specifically includes components of DOM that remain largely uncharacterized by conventional chromatographic analytical techniques targeting, for example, carbohydrates or amino acids. Linking the detailed information derived from non-targeted approaches in molecular microbiology and organic geochemistry has not been achieved to date, but it is a crucial first step toward a mechanistic understanding of the role of microorganisms in the global marine carbon cycle. In our study, we addressed this fundamental knowledge gap and connected the complex microbiological and molecular information obtained by ultrahigh-resolution methods via a novel combination of multivariate statistics.

Our study is based on samples from the North Sea, a 750 000 km<sup>2</sup> marginal sea of the Atlantic Ocean. The main water exchange with the Atlantic Ocean (51 000 km<sup>3</sup> year<sup>-1</sup> inflow and 56 700 km<sup>3</sup> year<sup>-1</sup> outflow) occurs via the northern boundary, while through the Dover Strait between the Netherlands and Great Britain ~4900 km<sup>3</sup> year<sup>-1</sup> enter the North Sea (Eisma and Kalf, 1987). Compared with these numbers, the flow into the system via precipitation, riverine freshwater input and the Baltic Sea in the East is relatively small and sums up to 895–955 km<sup>3</sup> year<sup>-1</sup> (Thomas *et al.*, 2005; Radach and Pätsch, 2007).

DOM, the dominant form of organic matter transported by rivers (Wetzel, 1984), is an important source of organic carbon and other nutrients (N, P) for coastal microbial metabolism, respiration and primary production (Azam *et al.*, 1983; Tranvik, 1992). The riverine DOM may either be derived from autochthonous primary production or it is of terrestrial origin (vascular plants and soils). From a molecular perspective, terrigenous components can be traced in aquatic systems by lignin-derived phenols (Opsahl and Benner, 1997; Hernes and Benner, 2006) and pyrogenic black carbon, a product of biomass burning (Dittmar and Paeng, 2009; Jaffé *et al.*, 2013). Mobilized from soils and subsequently transported to the oceans by rivers, terrigenous DOM contributes between <1% of the dissolved organic carbon (DOC) in the Pacific and up to 33% of the DOC in the Arctic Ocean (Opsahl and Benner, 1997; Opsahl *et al.*, 1999; Hernes and Benner, 2002). The decomposition pathways of terrigenous DOM in the ocean are still not fully constrained. Especially in coastal areas, which are the first receivers of terrigenous DOM, remineralization is a possible source of nutrients for primary production (Moran and Hodson, 1994; Opsahl and Benner, 1997; Fichot and Benner, 2014).

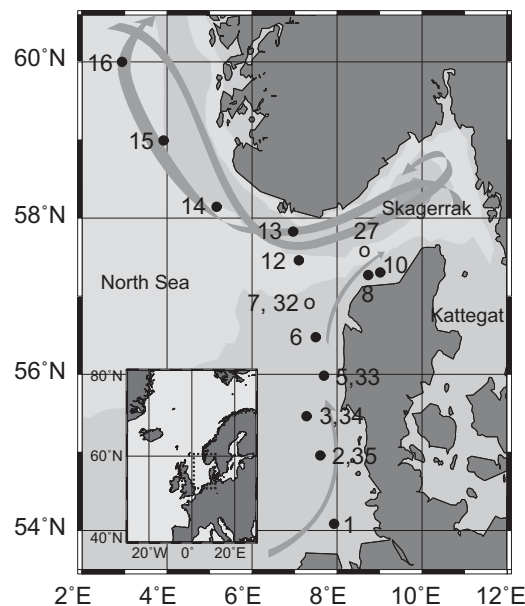
In this study, we molecularly characterized DOM at ultrahigh resolution via FT-ICR-MS, and statistically interpreted the molecular DOM data in the

context of the composition of the total and active BC, assessed by pyrosequencing of 16S ribosomal DNA (rDNA) and ribosomal RNA (rRNA). We aimed to (i) resolve the molecular composition of the North Sea DOM pool along a gradient of terrestrial influence starting in the German Bight and ending at approximately 60°N, and (ii) identify environmental variables as well as features of the BC associated with the gradient of DOM composition. We hypothesized that the BC composition is closely related to the composition of DOM, as it is the prevailing energy and carbon source for the largely heterotrophic bacteria.

## Materials and methods

### Study area and sampling

Water samples were obtained along a transect of 15 stations from the German coast toward Norway on R/V Heincke cruise HE361 from 12 July to 29 July 2011 (Figure 1). In addition, three stations (stations 8, 10 and 27) were sampled in the Skagerrak, the strait connecting the Baltic and the North Sea. At all stations, surface water was collected at 3–4 m depth using 5 l Niskin bottles attached to a CTD (Sea-Bird Electronics Inc., Bellevue, WA, USA); additional sampling depths were selected depending on salinity, water temperature and fluorescence (Wetlabs ECO FL(RT)) data reported by the CTD. In total, our study encompasses 45 DOM samples (Supplementary Tables S1). Chlorophyll *a* concentration was determined after extraction in 90%



**Figure 1** Sampling stations in the North Sea. Inset shows the respective North Atlantic area with the study area marked by a dashed rectangle. Circles denote sampling stations, filled circles indicate stations where RNA/DNA-based pyrosequencing data are available for surface samples and arrows indicate approximate directions of the main currents after Turrell (1992). Map created with Ocean Data View (R Schlitzer, <http://odv.awi.de>).

acetone according to Tremblay *et al.* (2002). For 12 of the samples for which DOM molecular information was available, the BC composition was analyzed by 454 pyrosequencing.

#### DOM analysis

Free and total dissolved neutral sugars were quantified spectrophotometrically after Mykkestad *et al.* (1997) and are given as glucose equivalents. Water samples for DOM analysis were filtered through GF/F filters (precombusted 400 °C, 4 h, Whatman, Maidstone, UK) and acidified to pH 2 with 25% HCl (p.a. grade, Carl Roth, Karlsruhe, Germany). Triplicate subsamples for DOC and total dissolved nitrogen (TDN) analysis were analyzed as described in Osterholz *et al.* (2014). The standard deviation of triplicate sample analyses was on average  $2 \mu\text{mol C l}^{-1}$  for DOC and  $1 \mu\text{mol N l}^{-1}$  for TDN analysis. According to the analysis of a deep Florida Strait reference (DSR) material (DA Hansell, University of Miami, FL, USA,  $n = 10$ ), the precision was  $2.8 \pm 0.9\%$  for DOC and  $3.3 \pm 1.1\%$  for TDN (computed as coefficient of variation across DSR runs), whereas the accuracy was  $5.1 \pm 4.2\%$  for DOC and  $-4.8 \pm 1.2\%$  for TDN (computed as average  $\pm$  s.d. of relative bias from reference concentrations), respectively. For mass spectrometry, DOM was solid-phase extracted from 2 l of filtered and acidified seawater with commercially available modified styrene divinyl benzene copolymer cartridges (PPL, Agilent, Santa Clara, CA, USA) as described in Dittmar *et al.* (2008). The extraction efficiency was on average  $52 \pm 5\%$  on a carbon basis.

Mass spectrometric analyses were performed on a 15 Tesla Solarix FT-ICR-MS (Bruker Daltonics, Billerica, MA, USA) as described in Seidel *et al.* (2014). Detected masses were matched across all samples, and molecular formulae (maximum elemental abundances set to  $\text{C}_n\text{H}_n\text{O}_n\text{N}_4\text{S}_1$ ) were assigned to masses with a minimum signal-to-noise ratio of 4 and allowing a mass error of  $<0.5$  p.p.m., using in-house programmed Matlab (MathWorks, Natwick, MA, USA) routines. Based on the elemental composition, molecules detected via FT-ICR-MS were assigned to 'molecular categories' (black carbon, polyphenols, highly unsaturated/unsaturated aliphatics, saturated fatty acids, carbohydrates, peptides) as proposed by Santl-Temkiv *et al.* (2013). Owing to the high number of possible structural isomers behind each molecular formulae (Hertkorn *et al.*, 2006), this classification is not unambiguous but exclusive (that is, categories are not overlapping) and it provides a useful, though rough, overview on likely molecular classes in DOM.

#### BC composition

Bacterioplankton were harvested from prefiltered 10 l water sample using a filter sandwich consisting of a glass fiber filter (142 mm diameter, GF/F,

Whatman) and 0.2  $\mu\text{m}$  polycarbonate filter (142 mm diameter, Nuclepore, Whatman). Prefiltration was performed according to Wemheuer *et al.* (2012). DNA and RNA were extracted from one-sixth of the filter sandwich (Weinbauer *et al.*, 2002). RNA was further purified and converted to complementary DNA (Wemheuer *et al.*, 2012). DNA was purified according to Wemheuer *et al.* (2015).

To assess BC structures, the V3–V5 region of the bacterial 16 S rRNA gene was amplified by PCR with the following set of primers (Muyzer *et al.*, 1995) containing the Roche 454-pyrosequencing adaptors, keys and one unique multiplex identifier per sample: 341 f 5'-CCATCTCATCCCTGCGTGTCTCCGAC-TCA G-(dn)<sub>10</sub>-CCTACGGGRAGGCAGCAG-3' and 907r 5'-CTATCCCCTGTGTGCCTTGGCAGTC-TCAG-CCGT CAATTCMTTGTGAGT-3'. The PCR reaction was performed as described in Wemheuer *et al.* (2012) with 50 ng of DNA or 25 ng of complementary DNA as template with one modification: the annealing temperature was 63 °C. Obtained PCR products were purified and quantified (Wemheuer *et al.*, 2015). The Göttingen Genomics Laboratory determined the sequences of the amplified PCR products using a Roche GS-FLX+ 454 pyrosequencer with Titanium chemistry (Roche, Mannheim, Germany).

Generated 16S rRNA gene and rRNA data sets were processed and further analyzed using the QIIME 1.8.0 software package (Caporaso *et al.*, 2010) and other tools. In brief, sequences were preprocessed and subsequently denoised using QIIME. Remaining reverse primer sequences and chimeras were removed (Wemheuer *et al.*, 2015). Processed sequences of all samples were joined and clustered in operational taxonomic units (OTUs) at 1%, 3% and 20% genetic distance (Wemheuer *et al.*, 2014). Taxonomy for each OTU at 1% genetic distance was determined by BLAST alignment against a modified version of the Silva database as outlined in Wemheuer *et al.* (2015). The modification concerned the reclassification of sequences affiliated the *Roseobacter* OCT lineage, which in fact consists of members of the CHAB-I-5, NAC1-2 and NAC11-6 lineages. Rarefaction curves and alpha diversity indices were calculated with QIIME as described by Wemheuer *et al.* (2013).

#### Statistical evaluation

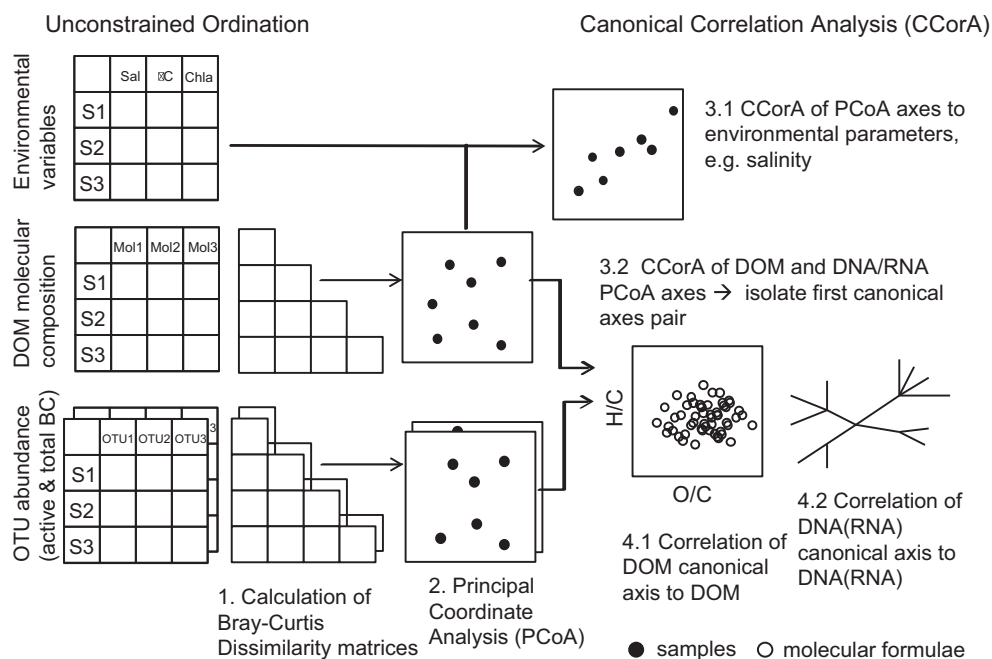
The BC composition was expressed in terms of relative abundances (as fractions of 1) of OTUs for each of the six phylogenetic resolutions from phylum to species. Similarly, the DOM composition was assessed based on the relative peak intensities of the FT-ICR-MS spectra, that is, absolute peak intensities normalized to the total sum of peak intensities of a given spectrum. It is important to emphasize here that FT-ICR-MS is a semiquantitative method. Concentrations of individual compounds in DOM cannot be determined because of the lack of calibration standards. Owing to differences in

ionization efficiency, the signal intensities of different masses are not necessarily proportional to their respective concentrations, but nevertheless increase/decrease in response to changes in concentration. We minimized potential matrix effects by adjusting the sample to identical concentrations in a solvent matrix of ultrapure water:methanol (LC-MS grade) at 1:1 ratio. Further, all samples considered in this study were highly similar in their molecular DOM composition, which facilitates reproducibility. Only masses with assigned molecular formula were taken into account for two data sets: DOM45 ( $n=6542$  masses) included all measured samples, DOM12 ( $n=6338$ ) included the DOM analyses at the 12 stations for which complementary pyrosequencing data were available.

A graphical overview of the steps taken in the statistical approach are summarized in Figure 2, the steps numbered in the figure are referred to in the following paragraph in brackets. To describe gradients of compositional change in the BC and in DOM, we used matrices of relative abundance (or intensity, respectively) to compute separate dissimilarity matrices based on Bray–Curtis compositional dissimilarity (Bray and Curtis, 1957; Legendre and Legendre, 1998; Figure 2.1). For each dissimilarity matrix, principal coordinate analysis (PCoA, aka metric scaling) was then used to graphically describe relationships among sites (that is, samples) based on the first two major axes of variation (Figure 2.2). The following canonical analyses targeted (i) the

relationships of BC or DOM composition to environmental factors and (ii) the linkage between BC and DOM compositions. The first was achieved by running a canonical correlation analysis (CCorA) using the BC-based or DOM-based PCoA axes and a chosen environmental factor (Singer *et al.*, 2012; Figure 2.3.1). This analysis aims at identifying a gradient of compositional change (a canonical axis in multivariate BC or DOM space), which most strongly correlates with the environmental factor. Also known as a canonical analysis of principal coordinates (Anderson and Willis, 2003), CCorA offers the possibility of a formal hypothesis test by permutation (Legendre *et al.*, 2011). We here note that CCorA of, for instance, DOM with an environmental variable may yield a significant result, whereas the same environmental variable is not necessarily significantly correlated with a first principal coordinate of DOM. This merely indicates that a relationship of DOM composition with the environmental variable must be rather subtle and does not express itself in major compositional changes of DOM; in such a case a canonical analysis (as CCorA) offers the necessary power to detect the relationship, which, however, may remain undetected in an unconstrained, descriptive analysis such as PCoA.

To link BC composition with the composition of DOM, we first selected a subset of principal coordinates for both involved dissimilarity matrices (one for BC, one for DOM, Figure 2.3.2) to avoid overfitting. This selection logically has to start with



**Figure 2** Flowchart demonstrating the statistical approach applied in the study. Detailed explanation is provided in the Materials and methods section of the manuscript. Unconstrained ordination (PCoA, 2) based on Bray-Curtis dissimilarity matrices (1) of DOM molecular composition and total/active microbial community composition (OTU abundances) was performed for dimension reduction. CCorA was used to find associations between environmental variables (3.1), DOM and BC composition (3.2). The associations were displayed as color-coded correlations in van Krevelen diagrams for the DOM molecular composition (4.1) and in unrooted trees for the microbial community (4.2).

the first PCoA axis, the one with the highest eigenvalue and thus covering the greatest share of compositional variation of the data set (Legendre and Legendre, 1998), and continue toward minor axes, with the aim to ideally lose no important information yet to also avoid too many dimensions contributing noise and hampering the analysis. This preselection is crucial, as the number of differentiated bacterial and DOM 'species' is much larger than the sample size, and a CCorA run on the original matrices would just result in a meaningless perfect fit. This step was done based on scree plots from PCoA output and with the aim to include at least 75% of the total variation of each data set in the canonical analysis step. We then fed the selected subsets of principal coordinates (PCoA1-3 for DOM and PCoA1-4 for BC, Supplementary Figure S1) into a CCorA aiming to identify the first canonical axis pair with maximized correlation between BC composition and the composition of DOM. The pair of correlated canonical axes consists of two linear combinations of the principal coordinates representing two specific dimensions in DOM and BC space. Significance of canonical correlation was computed by permutation ( $n=9999$ ), that is, sample allocation to sites of one used matrix was repeatedly randomized, the whole procedure until CCorA repeated, and the obtained null-distribution of such randomized canonical correlations used to compute significance of the observed canonical correlation (Rosenblad, 2009; Legendre *et al.*, 2011). Bimultivariate redundancy coefficients denote how much of the DOM composition is explained by the DNA- (or RNA-) based BC composition and vice versa. Note that CCorA is a symmetrical analysis (in the sense of a correlation), whereas bimultivariate redundancy coefficients are based on an asymmetric assessment of explanatory power involving an independent and a dependent matrix similar to classic regression.

To check whether the subset of principal coordinates was able to capture the relevant parts of variation contained in the full dissimilarity matrices, we added a Mantel test checking for a significant association of BC composition with the composition of DOM (Supplementary Table S4). This test is often viewed critically (Legendre and Fortin, 2010), yet it offers a global hypothesis test of association between dissimilarity matrices without information loss.

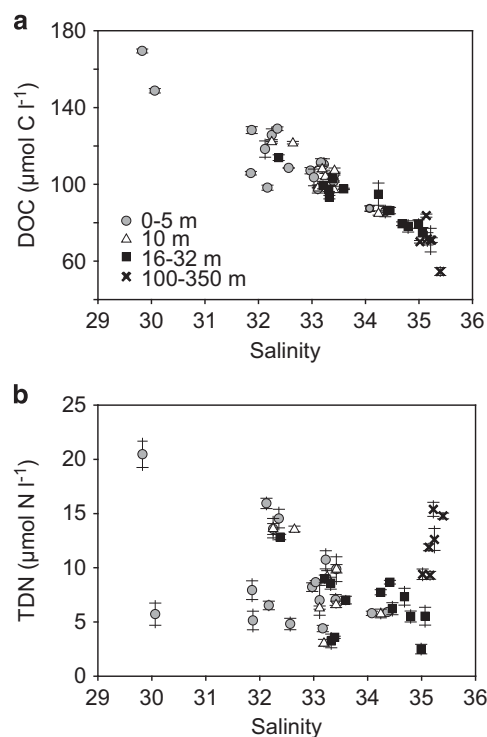
Our approach inevitably loses (expectedly irrelevant) information contained in minor principal coordinates, yet the canonical analysis has one major advantage: the association between BC and DOM identified in the form of a pair of canonical axes can be meaningfully related to the original multivariate matrices of BC and DOM, that is, the relative abundances of OTUs and intensities of formulae. To achieve this last step, we (i) correlated relative peak intensities with the canonical axis in DOM space and color-coded molecular formulae in Van Krevelen diagrams (Kim *et al.*, 2003) according to this correlation (Figure 2.4.1), and (ii) correlated

relative abundances with the canonical axis in BC space and color-coded OTU names on a phylogenetic tree according to this correlation (Figure 2.4.2). All multivariate statistics were performed in R (version 3.1.1, R Development Core Team, <http://cran.r-project.org/>) and using the package 'vegan' (Oksanen *et al.*, 2012).

## Results and discussion

### *Sources and distribution of DOC and TDN in the North Sea*

Salinity along our 700 km transect ranged from 29.8 at the southernmost station (station 1) to 35.4 at almost 60°N (station 16) with slight vertical stratification. Exceptions were the tidally impacted areas south of 56°N and the shallow Skagerrak (stations 8, 10 and 27), where no salinity stratification was observed. Our results reflect North Atlantic water flowing in at the northern boundary, circling through the North Sea (Thomas *et al.*, 2005) and mixing with riverine freshwater along the mainland coast and less saline Baltic Sea Water in the Skagerrak (Danielsen *et al.*, 1996). DOC concentrations ranged between 68 and 170  $\mu\text{mol C l}^{-1}$  and correlated significantly with salinity (Pearson's  $r=-0.94$ ,  $P<0.001$ ; Figure 3a). This finding indicates conservative mixing of high-DOC freshwater, largely consisting of riverine input from Elbe, Weser and Ems at the German coast, as well as Rhine, Meuse and Scheldt in the Netherlands

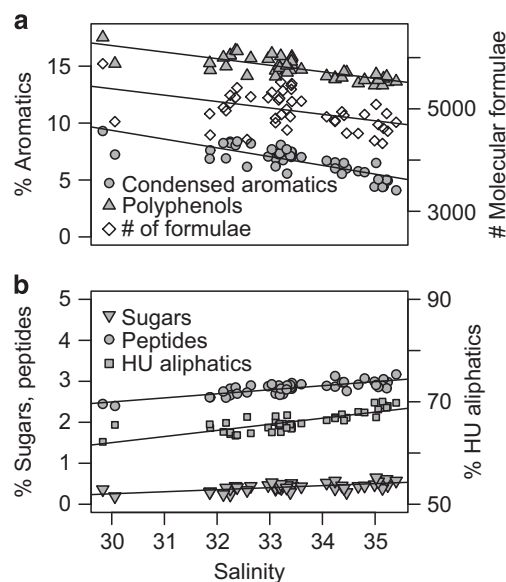


**Figure 3** (a) DOC and (b) TDN concentrations ( $\pm$  s.d. of triplicate analyses) versus salinity grouped by depth (0–5 m,  $n=18$ ; 10 m,  $n=8$ ; 16–32 m,  $n=13$ ; 100–350 m,  $n=6$ ) at all stations.

(Radach and Pätsch, 2007), and Atlantic water with lower DOC concentrations (Hansell and Carlson, 1998). According to Thomas *et al.* (2005), the low-salinity and high-DOC Baltic Sea inflow impacts the North Sea water salinity and DOC concentration in the area along a narrow strip off the Norwegian coast. In our study, this was merely reflected in the elevated DOC concentration of  $149 \mu\text{mol l}^{-1}$  at the low salinity (30.1) surface water of station 13 located above the Norwegian Trench. At the oceanic end of our investigated transect (station 16), DOC concentration was lowest ( $54 \mu\text{mol C l}^{-1}$ ). This is still higher than the concentrations reported for the Greenland Sea Deep Water ( $48 \mu\text{mol C l}^{-1}$ ) or the North Atlantic Deep Water ( $45 \mu\text{mol C l}^{-1}$ ; Hansell and Carlson, 1998), in agreement with the hypothesis by Thomas *et al.* (2003), according to which the North Sea acts as a 'bypass pump' that increases the carbon content of the Atlantic Ocean water during its circulation.

TDN concentrations ranged between 2.5 and  $20.5 \mu\text{mol N l}^{-1}$  (Figure 3b) with many surface stations being depleted in N, most likely due to removal during recent phytoplankton blooms (Colijn and Cadée, 2003). Water at depths between 100 and 350 m was enriched in TDN presumably due to the lack of active autotrophs and dominance of remineralization processes of sinking particles at these depths (Shaffer, 1996; Herndl and Reinthaler, 2013). During our sampling campaign, however, no extensive phytoplankton blooms were detected. Few stations exhibited chlorophyll maxima at 18–23 m depth, but chlorophyll *a* concentrations were generally low ( $<3.3 \mu\text{g l}^{-1}$ , mean  $1.3 \mu\text{g l}^{-1}$ ; Supplementary Table S2) when compared with the annual cycles shown in Radach and Pätsch (1997), where concentrations above  $10 \mu\text{g l}^{-1}$  were reached periodically.

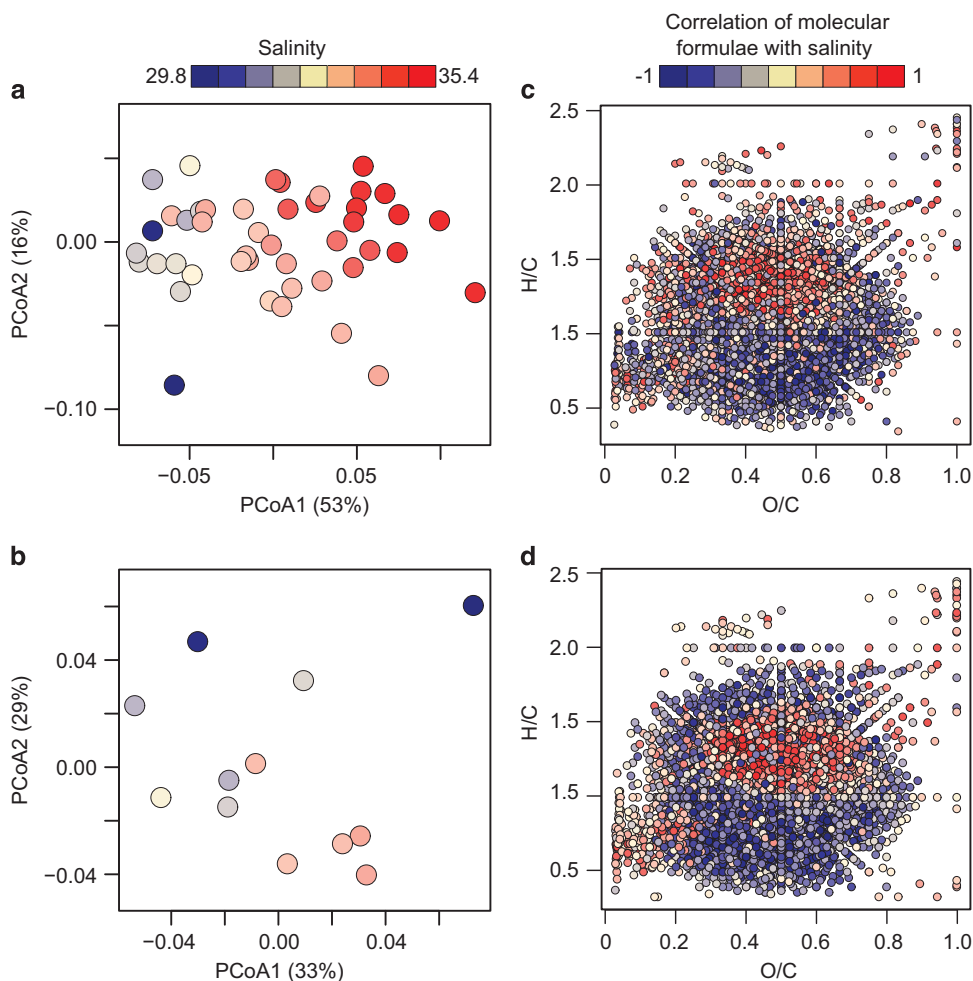
**Molecular DOM composition: tracing terrigenous input**  
Via FT-ICR-MS, we were able to resolve 10 198 different masses in the 45 DOM extracts. On average, molecular formulae were assigned to  $68 \pm 1\%$  of the masses, excluding  $^{13}\text{C}$  isotopologues. The average intensity-weighted molecular mass increased from 384 to 420 Da along the salinity gradient (Pearson's  $r=0.43$ ,  $P<0.05$ ). Accordingly, the relative contributions of molecular classes of DOM changed with salinity. Black carbon, a product of incomplete combustion (Goldberg, 1985; Jaffé *et al.*, 2013), and polyphenolic compounds, for example, humic substances (Opsahl and Benner, 1997), which are mainly derived from vascular plant material can act as, albeit not unambiguous, tracers of terrigenous DOM. The total number of molecular formulae, as well as the percentage of the total peak intensity of these two molecule classes, decreased significantly with increasing salinity (Figure 4). This indicates that the terrigenous DOM classes undergo similar conservative mixing as the bulk DOC concentration. In contrast, the fraction of highly unsaturated aliphatic molecules (Pearson's  $r=0.80$ ,  $P<0.005$ ),



**Figure 4** Relative proportion of DOM molecular groups: (a) percentages of condensed aromatic (black carbon) and polyphenol molecular formulae and total number of molecular formulae are negatively correlated to salinity, whereas (b) percentages of sugars, peptides and highly unsaturated (HU) molecular formulae increase with salinity. Black carbon formulae (Pearson's  $r=-0.80$ ), polyphenols ( $r=-0.77$ ), number of molecular formulae ( $r=-0.45$ ), sugars ( $r=0.65$ ), peptides ( $r=0.80$ ) and HU aliphatics ( $r=0.80$ ) all exhibit a significant ( $P<0.05$ ) correlation with salinity. Saturated fatty acid molecular formulae (not shown) did not vary with salinity.

peptides ( $r=0.80$ ,  $P<0.005$ ) and sugars ( $r=0.65$ ,  $P<0.05$ ) increased with increasing salinity, whereas the contributions of saturated fatty acid molecular formulae to the total formulae count remained constant. The highly unsaturated molecules defined here largely correspond to the carboxyl-rich alicyclic molecules described by Hertkorn *et al.* (2006): 87% of putative carboxyl-rich alicyclic molecule compounds defined through their DBE/C, DBE/O and DBE/H ratios fall into this group, another 11% are assigned to the polyphenol compound class.

PCoA including all 45 DOM samples resulted in a first principal coordinate, the axis of major compositional variation of DOM, which correlated significantly ( $P<0.001$ ) with salinity across the North Sea sampling transect (Figure 5a). Not surprisingly, canonical analysis of principal coordinates using salinity as a constraint revealed a highly significant ( $P<0.001$ ) association of DOM composition with salinity. This finding corresponds well to the previously described largely conservative mixing of the low-salinity coastal water masses with Atlantic Water in the North Sea and establishes salinity, the conservative mixing process itself, as the main determinant of DOM composition. However, based on the PCoA, 47% of the variability (covered by the second and subsequent PCoA axes) remained unexplained, pointing to substantial noise and/or additional factors that influence molecular DOM composition, for example, microbial interactions,



**Figure 5** PCoA based on Bray–Curtis dissimilarity of DOM molecular composition. Score plots (a) for all 45 samples and (b) for the reduced data set containing 12 samples, sites are color-coded by salinity. Van Krevelen plots show the correlations of molecular formulae (relative intensities) with salinity computed (c) from all 45 samples and (d) the reduced data set.

including production, degradation and transformation processes. As only 12 samples were available to investigate the relationship between DOM composition and BC composition, the statistical analysis was repeated with the reduced data set (Figure 5b). Salinity still explained 29% of the DOM compositional variability, but this salinity gradient was better expressed on the second PCoA axis. Again, CCorA revealed a significant association of salinity with DOM composition ( $P < 0.001$ ). In terms of fraction of explained variance, the first and second PCoA axes were nearly equivalent. Also, the weaker influence of salinity on the DOM composition may result from the fact that the reduced data set only included samples in a narrower salinity range (29.8–33.4). None of the measured environmental parameters such as chlorophyll *a* concentration, temperature, etc. could be confirmed as the main factor influencing DOM composition on the first PCoA axis, which accounted for 33% of the variability.

#### BC composition

After quality filtering, denoising and removal of potential chimeras and non-bacterial sequences, 336 367 high-quality sequences with an average read length of 524 bp (excluding primer sequences) were used to study total and active BC compositions. The number of sequences per sample ranged from 4470 to 19 847 (DNA level) and 1759 to 30 028 (RNA level). We were able to classify all sequences below phylum level. Calculated rarefaction curves, as well as diversity indices (data not shown), revealed that our surveying effort discovered the majority of—yet not all—the bacterial species and phyla (>80%) at 3% and 20% genetic distance. The active community yielded a higher number of OTUs than the total BC, resulting in a DNA/RNA ratio of OTU richness ranging from 0.76 to 0.88 and denoting that a significant portion of the rare taxa may be proportionally more active (Jones and Lennon, 2010; Campbell *et al.*, 2011; Wilhelm *et al.*, 2014). The percentage of OTUs that appeared in both, the RNA

**Table 1** Linking DOM composition to the composition of the total (DNA-based) and active (RNA-based) BC across six phylogenetic levels (phylum to species)

	<i>Phylum</i>	<i>Class</i>	<i>Order</i>	<i>Family</i>	<i>Genus</i>	<i>Species</i>
<i>(a) CanCor</i>						
DNA-DOM	0.95*	0.88	0.95*	0.95*	0.94**	0.94**
RNA-DOM	0.96*	0.96	0.93	0.96	0.97*	0.97*
<i>(b) Redundancy coefficients</i>						
DNAonDOM/DOMonDNA	0.27/0.54	0.34/0.38	0.47/0.63	0.48/0.64	0.53/0.65	0.53/0.66
RNAonDOM/DOMonRNA	0.30/0.40	0.28/0.48	0.36/0.52	0.42/0.46	0.52/0.58	0.50/0.61
<i>(c) Correlation of 1st canonical axes pair</i>						
DNA-DOM /RNA-DOM	0.80**	0.68*	0.20	0.54	-0.25	-0.52

Abbreviations: BC, bacterial community; CCorA, canonical correlation analysis; DOM, dissolved organic matter.

Given are (a) the canonical correlation coefficients resulting from CCorA (CanCor) and their significances (permutational probability, indicated with asterisks), and (b) bivariate redundancy coefficients for DOM composition as explained by DNA- and RNA-based BC composition (and vice versa). Also given are (c) the correlation coefficients (including significance as asterisks) of the first canonical axes pair. Asterisks denote significance levels.

Significance: \* $P < 0.05$ , \*\* $P < 0.01$ .

and DNA data set, decreased from 61% at phylum level to 40% at the highest phylogenetic resolution (Supplementary Table S3). A Mantel test confirmed strong correlations between DNA and RNA-based BC compositions at all phylogenetic levels ( $P < 0.01$ ; Supplementary Table S4), similar to a previous report on samples collected off the US-American east coast (Campbell *et al.*, 2011). Bacteria of the orders *Flavobacteriales* and *Rhodobacterales* were the most abundant taxa. Both phylogenetic groups are well known to be important during phytoplankton blooms (Teeling *et al.*, 2012; Buchan *et al.*, 2014). As shown by CCorA, the composition of the total BC at 'species' level (see reasons for restriction to this phylogenetic level in following paragraph) significantly correlated with chlorophyll *a* fluorescence (but not chlorophyll *a* concentration), DOC and TDN concentrations, whereas the active BC composition exhibited a significant correlation with TDN concentration only (Supplementary Table S5).

#### Linking BC and DOM across various levels of phylogenetic resolution

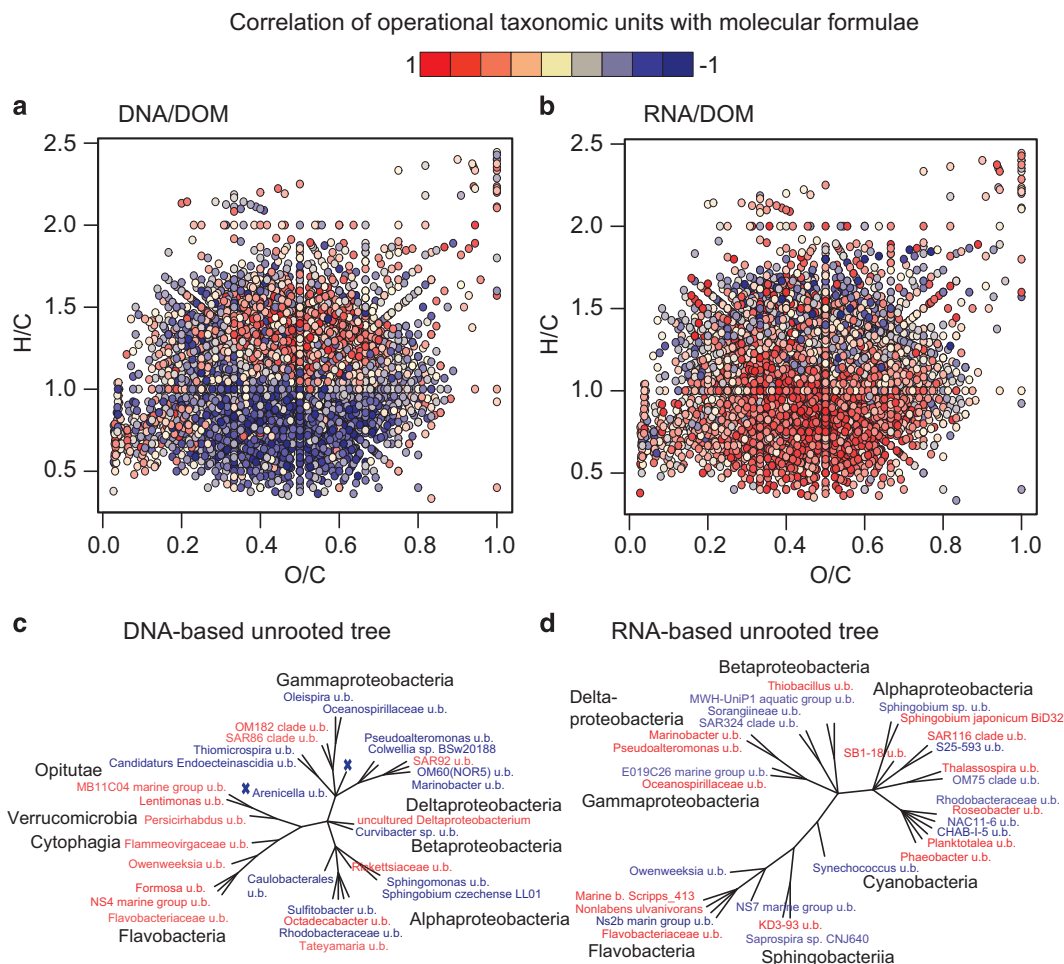
All statistical analyses involving the BC and DOM composition were performed separately for six levels of phylogenetic resolution from approximately phylum to species level (Supplementary Tables S4). CCorA analysis was chosen because the relationship between bacteria and DOM is bidirectional: bacteria may consume or produce DOM, but the prevalent DOM composition may also (and in turn) select for a certain microbial community. Applying this statistical tool, we detected a significant association between the total (DNA-based) BC and the molecular DOM composition at phylum and again at higher phylogenetic levels from order on. A significant association between the molecular DOM composition and the potentially active (RNA-based) BC composition was detected at phylum, genus and species levels, but not at other

phylogenetic levels. With higher phylogenetic level, the significance and the bivariate redundancy coefficients all exhibited (Table 1) constantly high values. However, the first canonical axes of the CCorA of DNA/DOM and RNA/DOM were correlated at the low resolution of phylum and class (Table 1): at these low phylogenetic levels the same phyla (or classes) of the active and total community correlate with DOM composition. This suggests that these levels of phylogenetic resolution are likely too coarse to differentiate the composition of the potentially active and the total community, let alone decipher differences in their relationships to DOM. Based on these results, we chose the 'species'-level for further analysis of relationships between the members of the active and of the total BC and specific DOM compounds (Figure 6).

#### Identifying the key factors—DOM molecules

The representation of the thousands of molecular formulae detected via FT-ICR-MS is simplified by plotting them according to their O/C versus H/C ratios in a van Krevelen diagram (Kim *et al.*, 2003; Figures 5 and 6). Most molecular formulae with positive correlations to the first canonical axes of salinity (Figures 5c and d) and DNA (Figure 6a) were relatively saturated (H/C ratio of  $> 1$ ), which is typical for marine DOM (Koch *et al.*, 2005; Sleighter and Hatcher, 2008; Kujawinski *et al.*, 2009; Medeiros *et al.*, 2015). Some combustion-derived condensed aromatic compounds ('black carbon', very low H/C and O/C ratios) also partially correlated with the canonical axis of DNA. The molecules that correlated positively with the canonical axes of salinity and DNA were on average larger (~430 Da) than the negatively correlated compounds (~350 Da). This finding corresponds well to a previous report stating that terrigenous DOM is on average of a smaller molecule size than the compounds characteristic for marine





**Figure 6** Van Krevelen plots show the correlations of molecular formulae (relative intensities) with the canonical axis associating DOM composition with (a) DNA-based BC composition and (b) RNA-based BC composition. Overall, 40% of molecular formulae occupy an identical location. Plotting was ordered by the absolute correlation value. Unrooted phylogenetic trees of OTUs on species level show the species with the 15 most positive (red) and the 15 most negative (blue) correlations with the canonical axis associating (c) DNA-based or (d) RNA-based BC composition with DOM (u.b. uncultured bacterium). Identical extreme colors of the color gradient (blue and red) for molecular formulae and OTUs indicate molecule–OTU associations, that is, red molecular formulae are associated with red OTUs, blue molecular formulae with blue OTUs.

environments (Koch *et al.*, 2005). Highly unsaturated compounds, including most polyphenols and some black carbon molecules, both typical compound classes of terrigenous DOM, correlated negatively with salinity and DNA. The average O/C ratio was similar for positively and negatively correlating compounds and ranged between 0.46 and 0.48, independent of whether the relationship with salinity or DNA was investigated.

Molecular formulae correlating with RNA showed a different distribution (Figure 6b). Compounds that negatively correlated with the RNA canonical axis were characterized by high H/C and variable O/C ratios. Some of these compounds contained nitrogen and are possibly smaller peptides, although other structural units with nitrogen cannot be ruled out. The numerous positively correlated molecules on the other hand spread over a wide range of H/C and O/C values.

It is important to emphasize that turnover rate and concentration of a substrate compound are often inversely related such as for dissolved carbohydrates (Simon and Rosenstock, 2007). A highly labile substrate to which bacteria have a high affinity can reach steady-state equilibrium concentrations below detection limit, although production and consumption rates are high. In contrast, more recalcitrant compounds can accumulate over longer time periods. It is therefore likely that the relationships between microbes and DOM compounds identified in this study relate mainly to semi-labile to more recalcitrant compounds, whereas highly labile compound groups like carbohydrates remained to some degree undetected.

In addition, solid-phase extraction of the samples with an efficiency of ~52% on a carbon basis may not have representatively captured the most labile components of the DOM such as amino acids,

peptides, mono- and polysaccharides (Hertkorn *et al.*, 2013; Osterholz *et al.*, 2015), which are rapidly turned over by the diverse, active BC (Simon and Rosenstock, 2007; Zubkov *et al.*, 2008). The applied electrospray ionization in negative mode potentially adds a further bias constraining our analytical window (Hertkorn *et al.*, 2008; Herlemann *et al.*, 2014). Nevertheless, our analysis did well in recovering fractions of highly unsaturated aliphatic molecules and some carbohydrates that both significantly increased with marine influence, and our statistics revealed that the abundance of some active OTUs (RNA-based) is related to potentially easily degradable compounds such as peptides (for example, Keil and Kirchman, 1999; Simon and Rosenstock, 2007; Figure 6b). Free and total dissolved neutral sugars were related to the salinity gradient, but did not show significant correlations to DOM, active or total community composition (Supplementary Table S5).

#### Identifying the key factors—BC

The 15 most positive and 15 most negative correlations of the DNA- and RNA-based OTUs exhibited correlation coefficients of  $r > 0.53$  and  $r < -0.47$ , respectively, with the DOM canonical axis 1 (Figures 6c and d). The ‘species’ level provided the highest canonical correlations for the DNA/DOM and the RNA/DOM analyses, which is a promising basis for further in-depth studies to elucidate functional and mechanistic relationships between these features. Transcriptomic or proteomic analyses of individual bacteria and communities (Teeling *et al.*, 2012; Ottesen *et al.*, 2014) alongside the identification of DOM compounds in the exometabolome (Romano *et al.*, 2014; Schwedt *et al.*, 2015) and marine DOM, prospectively on a structural level, are most suitable approaches for such investigations.

With regard to the total community (DNA-based), the OTUs correlating positively with the respective subgroup of DOM molecules were dominated by representatives of the *Proteobacteria* and *Bacteroidetes* phyla, especially the most abundant classes of *Gamma* (nine OTUs)—and *Alphaproteobacteria* (five OTUs) as well as *Flavobacteria* (Figure 6c). Negative correlations were also most abundant with members of *Gamma* (nine OTUs)—and *Alphaproteobacteria* (five OTUs), but all representatives of the *Bacteroidetes* and *Verrucomicrobia* phyla exhibited positive correlations. With regard to the active microbial community (RNA-based), the trends were less clear: representatives negatively correlating with the first canonical DOM axis included *Alphaproteobacteria* (seven OTUs) and *Flavobacteria* (three OTUs) and *Gammaproteobacteria* (three OTUs; Figure 6d); other *Alphaproteobacteria* (seven OTUs), *Flavobacteria* (three OTUs) and *Deltaproteobacteria* (two OTUs) correlated positively. These results, showing that bacteria of the same class can have opposing correlations to the DOM community,

emphasize the possible niche partitioning below class level facilitated by the vast DOM molecular space (Hutchinson, 1961; Gifford *et al.*, 2013).

#### Molecules and microbes—how do they relate?

Opposing views exist concerning the cycling and decomposition of highly diverse organic matter by marine bacterioplankton: a taxon-specific preference for specific DOM moieties was shown by Cottrell and Kirchman (2000), who found that *Cytophaga* and *Flavobacteria* dominated high molecular weight organic matter degradation processes. The results of Gomez-Consarnau *et al.* (2012) support the view that specific carbon compounds trigger the growth of certain bacterial strains, but also acknowledges the ‘generalists’—a functional redundancy of metabolic capabilities can lead to a minor influence of DOM quality on BC structure (Mou *et al.*, 2007). The approach taken in our study does not allow resolving any causal relationships between BC and DOM composition; rather, we generate new viewpoints and avenues for detailed research, which are catalyzed by the detailed insights into co-occurrences and absences between distinct molecular groups and bacterial lineages.

At species level, the canonical variates for the RNA/DOM and DNA/DOM analyses were not related, indicating that the association of the active community with DOM must be markedly different, that is, involving different OTUs, from the association of the total community with DOM. This is certainly a consequence of the increased compositional difference between the active and the total community at this high phylogenetic level, but it also underlines the complexity of the relationship between DOM and active versus total BC.

Overall, the correlation between the total BC and DOM composition is quite strong (Table 1) and seemingly driven by salinity. Indeed, a *post-hoc* analysis step identified the first canonical variate relating total BC and DOM composition as strongly correlated to salinity (Pearson’s  $r = -0.78$ ). Hence, even if DNA-based BC composition is individually not significantly correlated with salinity (Supplementary Table S5), the canonical correlation to DOM molecular composition identifies the connection to terrestrial influence as an important driver of the North Sea. The presence of a general salinity gradient mirrored in the BC composition has been reported previously, for example, by Bouvier and del Giorgio (2002) and Fortunato *et al.* (2012), who established salinity as the main driver of BC structure along a pronounced salinity gradient from river to ocean. In accordance to their findings, we hypothesize that the DOM compounds and the total BC that reach the North Sea undergo mixing and transformation processes from the coast toward the Atlantic on similar timescales of several months and thus retain the intermediate history of the water masses, the ‘riverine legacy’. Although taking place

on similar timescales, processes influencing DOM and total BC composition are probably different: for example, the BC may respond to changes in temperature or nutrient supply (Fuhrman *et al.*, 2008), whereas the DOM may be more affected by abiotic adsorption/desorption to surfaces (Gogou and Repeta, 2010), flocculation (Mikes *et al.*, 2004) or photochemical processes (Stubbins *et al.*, 2010).

OTUs only occurring in the DNA-based approach may include rare biosphere taxa that wait for more favorable growth conditions agreeing with their ecological niche (Pedrós-Alió, 2012; Müller *et al.*, 2014) or represent terrestrial or freshwater ‘leftovers’. Generally, the number of OTUs recovered in the DNA-based approach is approximately 20% lower compared with the RNA-based approach. Rare species suppressed in the analysis of the total BC can, however, be more active than the more abundant species (Campbell *et al.*, 2011). The general correlation between active BC composition and DOM was weaker than the DNA/DOM correlation (Table 1). One factor strengthening the correlation between total BC and DOM along the salinity gradient while diluting the active BC–DOM correlation may be the type of DOM that is affected: the terrigenous DOM driving the DNA/DOM relationship is only consumed by the BC (Fichot and Benner, 2014), whereas the more labile compounds affecting the active BC–DOM relationship are both consumed and produced. The aforementioned methodological constraints favoring the more refractory DOM compounds additionally weaken the active BC–DOM correlation. Together with the fact that the first canonical variate relating RNA-based BC and DOM was neither correlated to salinity nor to the DNA/DOM canonical variate, we propose that the correlations between the active BC and DOM were driven by ‘patchy’, short-lived events that are not very well covered by our survey effort. Such restricted events may include local and short-term phytoplankton blooms or confined particle-rich water masses caused by turbulent, wind-driven mixing that may influence microbial activity (Chrost and Riemann, 1994; Riemann *et al.*, 2000; Ziervogel and Arnosti, 2009). Owing to the multitude of possible local events in the highly dynamic North Sea (Sperling *et al.*, 2012), a much higher spatial and temporal sampling resolution is required to cover the natural variability of both DOM and microbial community composition to an extent sufficient for revealing general trends of prevailing DOM–microbe interactions. The rapid response of microbial communities to such changing conditions was recently demonstrated by Ottesen *et al.* (2014).

## Conclusion

This study examines the relationship of the DOM molecular composition and the structure of the active as well as the total microbial community in unprecedented detail. We were able to validate our

proposed advanced statistical approach by showing that the labile, rapidly turned over DOM is cycled by the active BC, constrained by local events such as phytoplankton blooms or storms. Contrary, driven by mixing processes, the (semi-) refractory DOM and total BC retain the history of the water masses, the riverine legacy, on similar timescales of several months. Most markedly, our results emphasize that the manifold interactions between microbes and molecules are undetectable at low analytical resolution. For instance, there were only weak to no relationships detected at the genus or family level, and, in analogy, at a compound group level. With state-of-the-art ultrahigh-resolution methods we were able to decipher interactions otherwise hidden through intrinsic averaging. The complete interactive network between molecules and organisms will only be visible with further enhancements in analytical resolution on a molecular structural and organismic functional level.

## Conflict of Interest

The authors declare no conflict of interest.

## Acknowledgements

We thank the captain, crew and scientific party of HE361 for making this cruise a success. We are grateful to Katrin Klapproth for assistance with processing FT-ICR-MS data and to Matthias Friebe for DOC and TDN analysis. Funding was provided by the University of Oldenburg (Germany) and DFG (Collaborative Research Centre TRR 51). Cruise oceanographic data (<http://dx.doi.org/10.1594/PANGAEA.785574>) and DOM molecular composition (<http://dx.doi.org/10.1594/PANGAEA.848886>) are available at PANGAEA. Sequence data were deposited in the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) under the accession number SRA082674.

## References

- Anderson MJ, Willis TJ. (2003). Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology* **84**: 511–525.
- Azam F, Fenchel T, Field JG, Gray JS, Meyerreil LA, Thingstad F. (1983). The ecological role of water-column microbes in the sea. *Mar Ecol-Prog Ser* **10**: 257–263.
- Bouvier TC, del Giorgio PA. (2002). Compositional changes in free-living bacterial communities along a salinity gradient in two temperate estuaries. *Limnol Oceanogr* **47**: 453–470.
- Bray JR, Curtis JT. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecol Monogr* **27**: 326–349.
- Buchan A, LeClerc GR, Gulvik CA, Gonzalez JM. (2014). Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nat Rev Microbiol* **12**: 686–698.

- Campbell BJ, Yu LY, Heidelberg JF, Kirchman DL. (2011). Activity of abundant and rare bacteria in a coastal ocean. *Proc Natl Acad Sci USA* **108**: 12776–12781.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.
- Chrost RJ, Riemann B. (1994). Storm-stimulated enzymatic decomposition of organic matter in benthic-pelagic coastal mesocosms. *Mar Ecol Prog Ser* **108**: 185–192.
- Colijn F, Cadée GC. (2003). Is phytoplankton growth in the Wadden Sea light or nitrogen limited? *J Sea Res* **49**: 83–93.
- Cottrell MT, Kirchman DL. (2000). Natural assemblages of marine proteobacteria and members of the Cytophaga-Flavobacter cluster consuming low- and high-molecular-weight dissolved organic matter. *Appl Environ Microbiol* **66**: 1692–1697.
- Danielssen DS, Svendsen E, Ostrowski M. (1996). Long-term hydrographic variation in the Skagerrak based on the section Torungen–Hirtshals. *ICES J Mar Sci* **53**: 917–925.
- Dittmar T, Koch B, Hertkorn N, Kattner G. (2008). A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnol Oceanogr-Meth* **6**: 230–235.
- Dittmar T, Paeng J. (2009). A heat-induced molecular signature in marine dissolved organic matter. *Nat Geosc* **2**: 175–179.
- Dittmar T, Stubbins A. (2014). 12.6—Dissolved organic matter in aquatic systems. In: Holland HD, Turekian KK (eds). *Treatise on Geochemistry*, 2nd edn. Elsevier: Oxford, pp 125–156.
- Edwards RA, Dinsdale EA. (2007). Marine environmental genomics: Unlocking the ocean's secrets. *Oceanography* **20**: 56–61.
- Eisma D, Kalf J. (1987). Dispersal, concentration and deposition of suspended matter in the North Sea. *J Geol Soc* **144**: 161–178.
- Fichot CG, Benner R. (2014). The fate of terrigenous dissolved organic carbon in a river-influenced ocean margin. *Global Biogeochem Cycles* **28**: 300–318.
- Fortunato CS, Herfort L, Zuber P, Baptista AM, Crump BC. (2012). Spatial variability overwhelms seasonal patterns in bacterioplankton communities across a river to ocean gradient. *ISME J* **6**: 554–563.
- Fuhrman JA, Cram JA, Needham DM. (2015). Marine microbial community dynamics and their ecological interpretation. *Nat Rev Microbiol* **13**: 133–146.
- Fuhrman JA, Steele JA, Hewson I, Schwalbach MS, Brown MV, Green JL et al. (2008). A latitudinal diversity gradient in planktonic marine bacteria. *Proc Natl Acad Sci USA* **105**: 7774–7778.
- Gifford SM, Sharma S, Booth M, Moran MA. (2013). Expression patterns reveal niche diversification in a marine microbial assemblage. *ISME J* **7**: 281–298.
- Gogou A, Repeta DJ. (2010). Particulate-dissolved transformations as a sink for semi-labile dissolved organic matter: Chemical characterization of high molecular weight dissolved and surface-active organic matter in seawater and in diatom cultures. *Mar Chem* **121**: 215–223.
- Goldberg ED. (1985). *Black Carbon in the Environment. Properties and Distribution*. John Wiley & Sons Inc.: New York, NY, USA.
- Gomez-Consarnau L, Lindh MV, Gasol JM, Pinhassi J. (2012). Structuring of bacterioplankton communities by specific dissolved organic carbon compounds. *Environ Microbiol* **14**: 2361–2378.
- Hansell DA, Carlson CA. (1998). Deep-ocean gradients in the concentration of dissolved organic carbon. *Nature* **395**: 263–266.
- Harvey HR, Dyda RY, Kirchman DL. (2006). Impact of DOM composition on bacterial lipids and community structure in estuaries. *Aquat Microb Ecol* **42**: 105–117.
- Herlemann DPR, Manecki M, Meeske C, Pollehne F, Labrenz M, Schulz-Bull D et al. (2014). Uncoupling of bacterial and terrigenous dissolved organic matter dynamics in decomposition experiments. *PLoS One* **9**: e93945–e93945.
- Herndl GJ, Reinthaler T. (2013). Microbial control of the dark end of the biological pump. *Nat Geosci* **6**: 718–724.
- Hernes PJ, Benner R. (2002). Transport and diagenesis of dissolved and particulate terrigenous organic matter in the North Pacific Ocean. *Deep-Sea Res Pt* **49**: 2119–2132.
- Hernes PJ, Benner R. (2006). Terrigenous organic matter sources and reactivity in the North Atlantic Ocean and a comparison to the Arctic and Pacific oceans. *Mar Chem* **100**: 66–79.
- Hertkorn N, Benner R, Frommberger M, Schmitt-Kopplin P, Witt M, Kaiser K et al. (2006). Characterization of a major refractory component of marine dissolved organic matter. *Geochim Cosmochim Acta* **70**: 2990–3010.
- Hertkorn N, Frommberger M, Witt M, Koch BP, Schmitt-Kopplin P, Perdue EM. (2008). Natural organic matter and the event horizon of mass spectrometry. *Anal Chem* **80**: 8908–8919.
- Hertkorn N, Harir M, Koch BP, Michalke B, Schmitt-Kopplin P. (2013). High-field NMR spectroscopy and FTICR mass spectrometry: powerful discovery tools for the molecular level characterization of marine dissolved organic matter. *Biogeosciences* **10**: 1583–1624.
- Hutchinson GE. (1961). The paradox of the plankton. *Am Nat* **95**: 137–145.
- Jaffé R, Ding Y, Niggemann J, Vähätalo AV, Stubbins A, Spencer RGM et al. (2013). Global charcoal mobilization from soils via dissolution and riverine transport to the oceans. *Science* **340**: 345–347.
- Jones SE, Lennon JT. (2010). Dormancy contributes to the maintenance of microbial diversity. *Proc Natl Acad Sci USA* **107**: 5881–5886.
- Keil RG, Kirchman DL. (1999). Utilization of dissolved protein and amino acids in the northern Sargasso Sea. *Aquat Microb Ecol* **18**: 293–300.
- Kim S, Kramer RW, Hatcher PG. (2003). Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van Krevelen diagram. *Anal Chem* **75**: 5336–5344.
- Koch BP, Witt MR, Engbrodt R, Dittmar T, Kattner G. (2005). Molecular formulae of marine and terrigenous dissolved organic matter detected by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Geochim Cosmochim Acta* **69**: 3299–3308.
- Kraft NJB, Cornwell WK, Webb CO, Ackerly DD. (2007). Trait evolution, community assembly, and the phylogenetic structure of ecological communities. *Am Nat* **170**: 271–283.

- Kujawinski EB. (2002). Electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry (ESI FT-ICR MS): characterization of complex environmental mixtures. *Environ Forensics* **3**: 207–216.
- Kujawinski EB, Longnecker K, Blough NV, Vecchio RD, Finlay L, Kitner JB *et al.* (2009). Identification of possible source markers in marine dissolved organic matter using ultrahigh resolution mass spectrometry. *Geochim Cosmochim Acta* **73**: 4384–4399.
- Legendre P, Fortin M-J. (2010). Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Mol Ecol Resour* **10**: 831–844.
- Legendre P, Legendre L. (1998). *Numerical Ecology*, 2nd edn. vol. 24. Elsevier Science B.V.: Amsterdam, The Netherlands.
- Legendre P, Oksanen J, ter Braak CJF. (2011). Testing the significance of canonical axes in redundancy analysis. *Methods Ecol Evol* **2**: 269–277.
- Liu L, Yang J, Yu X, Chen G, Yu Z. (2013). Patterns in the composition of microbial communities from a subtropical river: effects of environmental, spatial and temporal factors. *PLoS One* **8**: e81232.
- Medeiros PM, Seidel M, Ward ND, Carpenter EJ, Gomes HR, Niggemann J *et al.* (2015). Fate of the Amazon River dissolved organic matter in the tropical Atlantic Ocean. *Global Biogeochem Cycles* **29**: 677–690.
- Mikes D, Verney R, Lafite R, Belorgey M. (2004). Controlling factors in estuarine flocculation processes: experimental results with material from the Seine Estuary, Northwestern France. *J Coast Res* **SI 41**: 82–89.
- Moran MA, Hodson RE. (1994). Support of bacterioplankton production by dissolved humic substances from 3 marine environments. *Mar Ecol Prog Ser* **110**: 241–247.
- Mou XZ, Hodson RE, Moran MA. (2007). Bacterioplankton assemblages transforming dissolved organic compounds in coastal seawater. *Environ Microbiol* **9**: 2025–2037.
- Müller AL, de Rezende JR, Hubert CRJ, Kjeldsen KU, Lagkourvardos I, Berry D *et al.* (2014). Endospores of thermophilic bacteria as tracers of microbial dispersal by ocean currents. *ISME J* **8**: 1153–1165.
- Muyzer G, Teske A, Wirsén C, Jannasch H. (1995). Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Arch Microbiol* **164**: 165–172.
- Myklestad SM, Skånøy E, Hestmann S. (1997). A sensitive and rapid method for analysis of dissolved mono- and polysaccharides in seawater. *Mar Chem* **56**: 279–286.
- Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB *et al.* (2012). *vegan: Community Ecology Package*. R package version 2.0-4. Available from: <http://CRAN.R-project.org/package=vegan>.
- Opsahl S, Benner R. (1997). Distribution and cycling of terrigenous dissolved organic matter in the ocean. *Nature* **386**: 480–482.
- Opsahl S, Benner R, Amon RMW. (1999). Major flux of terrigenous dissolved organic matter through the Arctic Ocean. *Limnol Oceanogr* **44**: 2017–2023.
- Osterholz H, Dittmar T, Niggemann J. (2014). Molecular evidence for rapid dissolved organic matter turnover in Arctic fjords. *Mar Chem* **160**: 1–10.
- Osterholz H, Niggemann J, Giebel H-A, Simon M, Dittmar T. (2015). Inefficient microbial production of refractory dissolved organic matter in the ocean. *Nat Commun* **6**: 7422.
- Ottesen EA, Young CR, Gifford SM, Eppley JM, Marin R, Schuster SC *et al.* (2014). Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. *Science* **345**: 207–212.
- Pedros-Alió C. (2012). The rare bacterial biosphere. *Ann Rev Mar Sci* **4**: 449–466.
- Radach G, Pätsch J. (1997). Climatological annual cycles of nutrients and chlorophyll in the North Sea. *J Sea Res* **38**: 231–248.
- Radach G, Pätsch J. (2007). Variability of continental riverine freshwater and nutrient inputs into the North Sea for the years 1977–2000 and its consequences for the assessment of eutrophication. *Estuaries Coasts* **30**: 66–81.
- Riemann L, Steward GF, Azam F. (2000). Dynamics of bacterial community composition and activity during a mesocosm diatom bloom. *Appl Environ Microbiol* **66**: 578–587.
- Romano S, Dittmar T, Bondarev V, Weber RJM, Viant MR, Schulz-Vogt HN. (2014). Exo-metabolome of *Pseudovibrio* sp. FO-BEG1 analyzed by ultra-high resolution mass spectrometry and the effect of phosphate limitation. *PLoS One* **9**: e96038.
- Rosenblad A. (2009). B. F. J. Manly: randomization, bootstrap and Monte Carlo methods in biology, third edition. *Comput Stat* **24**: 371–372.
- Schwedt A, Seidel M, Dittmar T, Simon M, Bondarev V, Romano S *et al.* (2015). Substrate use of *Pseudovibrio* sp. growing in ultra-oligotrophic seawater. *PLoS One* **10**: e0121675.
- Seidel M, Beck M, Riedel T, Waska H, Suryaputra IGNA, Schnetger B *et al.* (2014). Biogeochemistry of dissolved organic matter in an anoxic intertidal creek bank. *Geochim Cosmochim Acta* **140**: 418–434.
- Shaffer G. (1996). Biogeochemical cycling in the global ocean.2. New production, Redfield ratios, and remineralization in the organic pump. *J Geophys Res-Oceans* **101**: 3723–3745.
- Simon M, Rosenstock B. (2007). Different coupling of dissolved amino acid, protein, and carbohydrate turnover to heterotrophic picoplankton production in the Southern Ocean in austral summer and fall. *Limnol Oceanogr* **52**: 85–95.
- Singer GA, Fasching C, Wilhelm L, Niggemann J, Steier P, Dittmar T *et al.* (2012). Biogeochemically diverse organic matter in Alpine glaciers and its downstream fate. *Nat Geosci* **5**: 710–714.
- Sleighter RL, Hatcher PG. (2008). Molecular characterization of dissolved organic matter (DOM) along a river to ocean transect of the lower Chesapeake Bay by ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Mar Chem* **110**: 140–152.
- Sperling M, Giebel H, Rink B, Grayek S, Staneva J, Stanev E *et al.* (2012). Differential effects of hydrographic and biogeochemical properties on the SAR11 clade and *Roseobacter* RCA cluster in the North Sea. *Aquat Microb Ecol* **67**: 25–34.
- Stegen JC, Lin X, Konopka AE, Fredrickson JK. (2012). Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J* **6**: 1653–1664.

- Stubbins A, Spencer RGM, Chen H, Hatcher PG, Mopper K, Hernes PJ *et al.* (2010). Illuminated darkness: molecular signatures of Congo River dissolved organic matter and its photochemical alteration as revealed by ultrahigh precision mass spectrometry. *Limnol Oceanogr* **55**: 1467–1477.
- Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM *et al.* (2012). Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* **336**: 608–611.
- Thomas H, Bozec Y, de Baar HJW, Elkalay K, Frankignoulle M, Schiettecatte LS *et al.* (2005). The carbon budget of the North Sea. *Biogeosciences* **2**: 87–96.
- Thomas H, Gattuso JP, Smith SV. (2003). Coastal Biogeochemistry. EGS-AGU-EUG Joint Assembly.
- Tranvik LJ. (1992). Allochthonous dissolved organic matter as an energy source for pelagic bacteria and the concept of the microbial loop. *Hydrobiologia* **229**: 107–114.
- Tremblay JE, Lucas MI, Kattner G, Pollard R, Strass VH, Bathmann U *et al.* (2002). Significance of the polar frontal zone for large-sized diatoms and new production during summer in the Atlantic sector of the Southern Ocean. *Deep-Sea Res Pt II* **49**: 3793–3811.
- Turrell WR. (1992). New hypotheses concerning the circulation of the northern North Sea and its relation to North Sea fish stock recruitment. *ICES J Mar Sci* **49**: 107–123.
- Valentin-Vargas A, Toro-Labrador G, Massol-Deya AA. (2012). Bacterial community dynamics in full-scale activated sludge bioreactors: Operational and ecological factors driving community assembly and performance. *PLoS One* **7**: e42524.
- Weinbauer MG, Fritz I, Wenderoth DF, Hofle MG. (2002). Simultaneous extraction from bacterioplankton of total RNA and DNA suitable for quantitative structure and function analyses. *Appl Environ Microbiol* **68**: 1082–1087.
- Wemheuer B, Gullert S, Billerbeck S, Giebel HA, Voget S, Simon M *et al.* (2014). Impact of a phytoplankton bloom on the diversity of the active bacterial community in the southern North Sea as revealed by metatranscriptomic approaches. *FEMS Microbiol Ecol* **87**: 378–389.
- Wemheuer B, Taube R, Akyol P, Wemheuer F, Daniel R. (2013). Microbial diversity and biochemical potential encoded by thermal spring metagenomes derived from the Kamchatka peninsula. *Archaea* **2013**: 136714.
- Wemheuer B, Wemheuer F, Daniel R. (2012). RNA-based assessment of diversity and composition of active archaeal communities in the German Bight. *Archaea* **2012**: 695826.
- Wemheuer B, Wemheuer F, Hollensteiner J, Meyer F-D, Voget S, Daniel R. (2015). The green impact: bacterioplankton response towards a phytoplankton spring bloom in the southern North Sea assessed by comparative metagenomic and metatranscriptomic approaches. *Front Microbiol* **6**: 805.
- Wetzel RG. (1984). Detrital dissolved and particulate organic carbon functions in aquatic ecosystems. *Bull Mar Sci* **35**: 503–509.
- Wilhelm L, Besemer K, Fasching C, Urich T, Singer GA, Quince C *et al.* (2014). Rare but active taxa contribute to community dynamics of benthic biofilms in glacier-fed streams. *Environ Microbiol* **16**: 2514–2524.
- Ziervogel K, Arnosti C. (2009). Enzyme activities in the Delaware Estuary affected by elevated suspended sediment load. *Estuar Coast Shelf S* **84**: 253–258.
- Zinger L, Amaral-Zettler LA, Fuhrman JA, Horner-Devine MC, Huse SM, Welch DBM *et al.* (2011). Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. *PLoS One* **6**: e24570.
- Zubkov MV, Tarran GA, Mary I, Fuchs BM. (2008). Differential microbial uptake of dissolved amino acids and amino sugars in surface waters of the Atlantic Ocean. *J Plankton Res* **30**: 211–220.
- Šantl-Temkiv T, Finster K, Dittmar T, Hansen BM, Thyraug R, Nielsen NW *et al.* (2013). Hailstones: a window into the microbial and chemical inventory of a storm cloud. *PLoS One* **8**: e53550.

Supplementary Information accompanies this paper on The ISME Journal website (<http://www.nature.com/ismej>)