<https://helda.helsinki.fi>

Coupling biogeochemical process rates and metagenomic blueprints of coastal bacterial assemblages in the context of environmental change

Markussen, Trine

2018-08-01

Markussen , T , Happel , E M , Teikari , J E , Huchaiah , V , Alneberg , J , Andersson , A F , Sivonen , A K , Riemann , L , Middelboe , M & Kisand , V 2018 , ' Coupling biogeochemical process rates and metagenomic blueprints of coastal bacterial assemblages in the context of environmental change ' , Environmental Microbiology , vol. 20 , no. 8 , pp. 3083-3099 . <https://doi.org/10.1111/1462-2920.14371>

<http://hdl.handle.net/10138/324405>

<https://doi.org/10.1111/1462-2920.14371>

acceptedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Coupling biogeochemical process rates and metagenomic blueprints of coastal bacterial assemblages in the context of environmental change

Trine Markussen¹, Elisabeth M. Happel¹, Jonna E. Teikari², Vimala Huchaiiah³, Johannes Alneberg⁴, Anders F. Andersson⁴, Kaarina Sivonen², Lasse Riemann¹, Mathias Middelboe^{1*}, Veljo Kisand^{3*}

¹Marine Biological Section, Department of Biology, University of Copenhagen, Helsingør, Denmark

²Department of Microbiology, University of Helsinki, Helsinki, Finland

³Institute of Technology, University of Tartu, Tartu, Estonia

⁴KTH Royal Institute of Technology, Science for Life Laboratory, School of Biotechnology, Stockholm, Sweden

Running title: Bacterioplankton responses to perturbation

*Correspondence: Mathias Middelboe: mmiddelboe@bio.ku.dk, Veljo Kisand: kisand@ut.ee

Summary

Bacteria are major drivers of biogeochemical nutrient cycles and energy fluxes in marine environments, yet how bacterial communities respond to environmental change is not well known. Metagenomes allow examination of genetic responses of the entire microbial community to environmental change. However, it is challenging to link metagenomes directly to biogeochemical process rates. Here, we investigate metagenomic responses in natural bacterioplankton communities to selected environmental stressors in the Baltic Sea, including increased river input, increased nutrient concentration, and reduced oxygen level. This allowed us to identify informative prokaryotic gene markers, responding to environmental perturbation. Our results demonstrate that metagenomic and metabolic changes in bacterial communities in response to environmental stressors is influenced both by the initial community composition and by the biogeochemical factors shaping the functional response. Furthermore, the different sources of dissolved organic matter (DOM) had the largest impact on metagenomic blueprint. Most prominently, changes in DOM loads

26 influenced specific transporter types reflecting the substrate availability and DOC assimilation and
27 consumption pathways. The results provide new knowledge for developing models of ecosystem structure
28 and biogeochemical cycling in future climate change scenarios and advance our exploration of the potential
29 use of marine microorganisms as markers for environmental conditions.

30

31 **Introduction**

32 Prokaryotic microbial communities respond rapidly to environmental changes e.g. substrate composition,
33 oxygen concentration, nutrient supply and salinity by changes in their community composition and gene
34 expression (Kirchman *et al.*, 2004; Sjöstedt *et al.*, 2012; Lindh *et al.*, 2015). Consequently, the genetic
35 composition of prokaryotic microbial communities is linked to concomitant environmental conditions, and
36 changes in their combined gene-pool (genetic blueprint) may potentially be used as an indicator of
37 environmental changes. At the same time, prokaryotes are the main drivers of biogeochemical cycles, and
38 changes in their abundance, activity and metabolism may therefore affect the prevalence of specific
39 biogeochemical pathways (Fuhrman *et al.*, 2015; Graham *et al.*, 2016). These properties make prokaryotes
40 an important, but so far largely overlooked group of organisms to include in environmental monitoring
41 (Caruso *et al.*, 2016). Thus, identification of informative prokaryotic gene markers changing in abundance in
42 response to external forcing would be highly valuable in a monitoring context. Today, next generation
43 sequencing techniques applied to environmental samples (e.g. metagenomics and metatranscriptomics)
44 provide comprehensive and quantitative descriptions of the phylogenetic and functional diversity by using
45 the total pool of nucleic acids (Gilbert and Dupont, 2011; Anantharaman *et al.*, 2016). Collectively,
46 metagenomes and metatranscriptomes can be used to identify couplings between environmental conditions
47 and the microbial community composition and function. Likewise, recent studies used changes in
48 metagenomes to assess the impact of anthropogenic stressors on microbial communities (Kisand *et al.*, 2012;
49 Won *et al.*, 2017). Yet, linking such changes in the microbial blueprint with specific changes in

biogeochemical processes still remains a major challenge.

Our understanding of microbial diversity in the Baltic Sea (BS) was recently expanded by studies on bacterial biogeography (Herlemann *et al.*, 2011) and seasonal community succession (Andersson *et al.*, 2009; Herlemann *et al.*, 2013). Further, the identification of key prokaryotic players and their functional traits in subsystems of the BS, e.g. surface waters and the oxic-anoxic interfaces of the deep basins (Labrenz *et al.*, 2010; Grote *et al.*, 2011; Farnelid *et al.*, 2013), has made substantial progress in recent years. Recently, metagenomes from a time-series of sampling station in the central BS and cruises across the various basins were successfully assembled and used to establish a reference database (BARM – Baltic reference metagenome) (Alneberg *et al.*, submitted), enabling analysis of any BS metagenome - or transcriptome sample by mapping the sequences onto the database and deducing the functional and/or taxonomic profile of the community.

In this study, we explored metagenomic blueprints of natural microbial assemblages exposed to selected stressors known to constitute major environmental concerns in the BS (e.g. reduced oxygen, cyanobacterial blooms, and increased input of riverine water). By performing experiments using seawater microcosms (10 L) from two sites in the BS representing different sub-basins, we sought to link these blueprints to microbial abundance, productivity and activity measurements in order to quantify changes in the microbial metagenomes and activities in response to environmental perturbations. We linked effects of anthropogenic disturbances on bacterial communities and biogeochemical processes with the abundance of specific protein clusters and identified important DOC-transporting proteins in the metagenomic blueprint that were affected by specific stressors. Our analysis adds new insights to the complex influence of environmental changes on microbial community composition and function and emphasizes the challenge of identifying specific signature genes to be used as indicators of environmental status in marine monitoring.

72

73 **Results**

74 *Characteristics of the investigated sites*

The two investigated sites in the BS (Fig. S1) were characterized by the nutrient concentration, salinity and temperature (Table 1). The dissolved organic carbon (DOC) concentration was nearly 2 fold lower at Øresund site (Exp I) compared to Storfjärden site (Exp II), in accordance with the large input of humic DOC from terrestrial sources to the Gulf of Finland. As the dissolved organic nitrogen (DON) concentrations were similar at the two stations, the C:N ratios of the organic matter differed substantially between Øresund (C:N = 7.6) and Storfjärden (C:N = 13.3). Inorganic nutrient concentrations were generally low at both sites but with a higher inorganic N:P ratio at Øresund (N:P = 23.3) than at the Storfjärden site (N:P = 1.7). Bacterial abundance (BA) and production (BP) was 1.6-2 fold lower at the Øresund site in April compared to the Storfjärden site in July. The same trend was seen for viral abundances (VA) (Table 1).

Abundance, production and activity of the bacterial community and viruses

In both experiments, an increase in cell abundance and production of the diluted natural bacterial communities was observed in all six treatments (DOMhum, addition of humic rich river water; DOMagri, addition of agricultural river water; DOMsea, addition of autochthonous plankton lysate; LTC, addition of toxic cyanobacterial cell lysate (*Nodularia spumigena*); NP, inorganic nutrient addition; LowO₂, reduced O₂ concentration), and in controls (Fig. 1). DOMagri treatment caused elevated initial VA in Exp II, reflecting a high viral concentration in that river (Fig. 1). The VA remained relatively constant during the incubation except for the LTC treatment in Exp I, where the increased BP supported a doubling of the VA during the stationary phase of the incubation.

Most pronounced bacterial growth was observed in the LTC treatment reflecting a large input of labile DOC in both experiments (Fig. 1 and S2). LTC treatments showed significant increase of protease (both experiments), β -glucosidase (both experiments) and alkaline phosphatase activities (Exp I) compared to controls (Fig. 2, Dunnett's test; Table S1). DOMagri treatments significantly increased BA (both experiments) and BP (Exp II), VA (both experiments), protease (both experiments), β -glucosidase (Exp II) and alkaline phosphatase activities (both experiments, Fig. 1 and 2) compared with the controls (Table S1). DOMhum treatment had no effect on bacterial growth in Exp I, but significantly stimulated BA and

101 production in Exp II (Table S1). DOMsea treatment significantly increased BA (Exp I), BP (Exp II), protease
102 activity (Exp I) and alkaline phosphatase activity (both experiments) compared to the controls (Fig. 1 and 2,
103 Table S1). Treatments with NP and LowO₂ did not affect BA or BP relative to the controls (both
104 experiments, Table S1).

105 The dynamics of inorganic nutrient concentrations differed between the two experiments, but also between
106 treatments (Fig. S3). DOMagri treatment constituted a significant input of NO₃ in Exp I but not in Exp II
107 (Table S1). The high DOC:DON and low inorganic N:P ratios at the Exp II site was associated with a
108 significant decrease in ammonium concentration in the treatments amended with inorganic N (DOMagri,
109 DOMhum and NP). This was not the case at the Exp I site (high N:P ratio), where the phosphate
110 concentration was reduced to below detection limit within 40 h.

111

112 *Community composition*

113 At the end of experiments (T_{end}) microbial community composition differed between Exp I and Exp II and
114 higher phylogenetic variability was observed in Exp II compared to Exp I (Fig. 3A and B). *Actinobacteria*
115 were more abundant in Exp II compared to Exp I (median Exp I: 0.7%; median Exp II: 3.3%) whereas
116 *Gammaproteobacteria* were both the most dominant sub-phylum in Exp I (median Exp I: 43.6%; median
117 Exp II 9.4%). In both experiments, the abundance of *Betaproteobacteria* was stimulated in the DOMhum
118 and DOMsea treatments relative to the controls. Addition of LTC changed the abundance of several sub-
119 phyla compared with controls including a large increase in *Gammaproteobacteria* and a decrease in
120 *Actinobacteria* and *Betaproteobacteria*.

121 The taxonomic richness (calculated as abundance-based coverage estimator (ACE)) was significantly higher
122 in Exp II compared to Exp I ($R^2 = 0.24$, $p < 0.001$), however, no significant differences in richness were
123 observed between treatments from same experiment (Fig. S4A). Furthermore, Shannon diversity was
124 significantly higher in bacterial communities for Exp II than Exp I ($R^2 = 0.67$, $p < 0.001$, Fig. S4B) and only
125 LTC bacterial communities (Exp I and II) showed significantly lower diversity compared to controls ($p <$
126 0.05).

127

128 *Functional metagenomic profile*

129 Metagenomic sequencing of the total community DNA from T_{end} resulted in 2.6 to 19.1×10^6 raw reads in
 130 Exp I and 5.0 to 12.4×10^6 raw reads in Exp II per sample. The number of reads that mapped to the BARM
 131 database varied between 1.1 and 3.9×10^6 for Exp I samples, and between 1.2 and 3.5×10^6 for Exp II
 132 samples, corresponding to 30 – 70% of the total number reads. Mapped reads were annotated using Cluster
 133 of Orthologous Groups (COGs).

134 Multidimensional scaling (MDS) plot of functional community profiles based on distances corresponding to
 135 log₂-fold-changes (logFC) between each sample showed that the treatments clustered according to
 136 experimental site (Fig. S5). However, COGs associated with distinct metabolic pathways varied between
 137 locations. For example, we observed that Exp I had higher abundance of the key enzyme 2-keto-3-deoxy-6-
 138 phosphogluconate aldolase (COG0800) in the Entner–Doudoroff (ED) pathway than Exp II ($p < 0.01$, logFC
 139 = 1.29), while Exp II treatments had higher abundance of the key enzyme 6-phosphofructokinase (COG0205,
 140 $p < 0.01$, logFC = 0.69) involved in the common glycolysis (Embden–Meyerhof–Parnas - EMP) pathway
 141 (data not shown).

142

143 *Effects of treatment on metagenomic blueprints*

144 In order to identify COGs which were representative for specific treatments, we compared treatments with
 145 controls from the given site. The number of significantly responding genes was higher for treatments in Exp
 146 II than in Exp I ($p < 0.01$, FDR > 0.05, Table 2). Especially, the LTC treatments affected a high number of
 147 COGs (72% of annotated COGs). In order to shortlist the number of COGs for more detailed analysis of
 148 metabolic pathways, an additional logFC filter was therefore applied for the statistical analysis of selected
 149 treatments (see *Experimental procedures*, Table 2) resulting in a total of 468 and 592 responding COGs
 150 selected for further analysis for Exp I and Exp II, respectively. The up-represented COGs accounted for 55%
 151 and 35% of total responding COGs for Exp I and Exp II, respectively (Table S2).

152 In both experiments we observed an overall dominance of responding COGs related to metabolism and

transport (Fig. S6) with DOC-related COGs accounting for 18-65% (Exp I) and 30-52% (Exp II) of all up
 represented COGs in the different treatments (Table 2, Table S3). In addition, some of the most up-
 represented COGs relative to the control treatments were from the categories 'Transcription' and 'Signal
 transduction mechanisms' with functions related to iron metabolism (COG1321), phage transcription
 (COG3311), sporulation (COG1774, COG3854), denitrification or amino acid/polyamine metabolism and
 transport (COG1348, COG4566, COG3284, COG2909) (Table S2). Further, in several treatments
 (DOMhum, DOMagri and NP in Exp I, and DOMhum and LowO₂ in Exp II), a relatively large fraction of
 the up represented COGs belonged to the 'Mobilome: prophages, transposons' category (Fig. S6).
 Cluster analysis of up-represented COGs unique to or shared between treatments demonstrated treatment-
 related division of functional properties (Fig. 3C and D). A k-medoids clustering analysis to partition the
 COGs into five clusters based on similarities in abundance profiles, identified distinct clusters of COGs
 associated with specific environmental stressors (Fig. 3E and F, Table S2). For example, cluster 4 were
 dominated by COGs from the DOMhum treatment (Exp I), and COGs from cluster 5 (Exp I) were only up-
 represented in the LowO₂ treatment (Fig. 3E, Table S2). The clustering further showed functional similarities
 between treatments, such as COGs in Cluster 3 (Exp II) which were up-represented in DOMhum, DOMagri
 and NP treatments (Fig. 3F, Table S2).
 Distinct categories of DOC-related COGs were affected in the different treatments (Table S3). In general, the
 most up-represented category in Exp I was 'Carbohydrate transport and metabolism' (19% of all up
 represented DOC-related COGs). Especially in the DOMhum, DOMsea and LTC treatments, the COG
 category 'Carbohydrate transport and metabolism' was highly up-represented relative to the control. In
 DOMhum, this was combined with COGs involved in phosphate transport, whereas in DOMsea, also COGs
 assigned to 'Inorganic ion and transport metabolism' (23%) and proteases responded to the treatment (Table
 S3). A large variety of COGs involved in carbohydrate metabolism was up-represented in the LTC treatment,
 including COGs associated with the metabolism of simple sugars (e.g. fructose, sucrose, mannose and
 starch). The metagenomic response for NP and DOMagri treatments was highly similar, sharing 71% of all
 up-represented COGs and with a limited number of up-represented DOC-related COGs (n = 4 for DOMagri

and n = 3 for NP). Likewise, no dominant COG category was affected by the LowO₂ treatment and the affected DOC-related COGs (n = 4) showed less than 0.7 logFC difference in abundance compared to controls (Table S3).

In contrast to Exp I, the most affected category in Exp II was 'Inorganic ion and transport metabolism' (17% of all up-represented DOC-related COGs. The DOMhum treatment in Exp II affected COGs involved in nucleotide and amino acid metabolism pathways (23%) rather than carbohydrate metabolism (12%). The response to DOMagri treatment in Exp II was also different from Exp I, showing high abundance of DOC-related COGs in the category 'Nucleotide and metabolism' (24%) and with 60 % overlap with DOMhum in responding COGs (Fig. 3D). Both treatments showed up-representation of enzymes associated with nitrogen metabolism (Table S3). As in Experiment I, COGs associated with carbohydrate metabolism was highly up-represented in the DOMsea and LTC treatment (Table S3). In treatments amended with various levels of inorganic N (DOMagri, DOMhum and NP) up-representation of COGs related to purine degradation was observed. A limited number of affected DOC-related COGs changed in abundance in LowO₂ (n = 3).

Comparison of the treatment-specific metagenomes between the two experiments identified a common set of COGs that responded similar to the imposed treatments, thus representing potential indicator COGs for the specific stressors (Fig. S7). Such common signature COGs with higher relative abundance compared to controls were particular prevalent in DOMhum and LTC treatments.

Shared COGs between DOMhum treatments involved COGs belonging to the 'Mobilome: prophages, transposons' category and five COGs involved in carbohydrate metabolism including three tripartite ATP-independent periplasmic (TRAP) transporters. Shared COGs between LTC treatments (logFC > 1.5) included two COGs assigned as ABC transporters, and COGs involved in carbohydrate metabolism and amino acid metabolism. The shared COGs in the DOMsea treatments also included DOC-related COGs (Fig. S7). No COGs were shared between LowO₂ treatments.

Changes in abundance of transporters

Changes in abundance of specific transporter genes were analyzed in the subset of significantly up-

205 represented COGs ($\log_{2}FC > 0$, $FDR < 0.05$, p -adjusted < 0.001 , see *Experimental Procedures*). In Exp I a
 206 number of important transporters responded to different treatments (Fig. 4). Both in the DOMhum treatment
 207 (ABC type and three TRAP type transporters) and in the DOMsea treatment (tonB-dependent transporters)
 208 specific transporters increased significantly in abundance compared with the controls. No transporters
 209 increased in abundance in DOMagri, NP and LowO₂ treatments compared to control, whereas many
 210 transporters were identified in LTC treatment, including one ABC-type transporter, seven permeases, one
 211 antiporter, two tonB-dependent transporters and three symporters. In total 23 transporter genes were
 212 increasing in abundance in response to the LTC treatment.
 213 The diversity of transporters was higher in Exp II (Fig. 4) compared to Exp I, with the most pronounced
 214 response in treatments receiving river water: 33 transporters responded with increasing counts in DOMhum
 215 treatment, including 19 ABC-transporters and five TRAP-transporters, and 13 transporters in DOMagri,
 216 including three ABC-transporters. LTC treatment was not dominated by any specific type of transporters (n
 217 = 4).
 218

219 *Linking activity and function: relationship between COG abundance and microbial activity*

220 Detrended correspondence analysis (DCA) of COG abundances responding significantly to the treatments
 221 and linear fitting of response variables identified significant correlations ($p < 0.05$) between cluster of COGs
 222 and activity/abundance/concentration variables. The linear fitting indicated that elevated BP, extracellular
 223 enzymatic activities and amino acids uptake were strongly associated to LTC treatment (data not shown)
 224 suggesting that the strong response from adding high loads of DOC from cyanobacterial lysate shaded the
 225 response from the remaining treatments. In order to analyze the relationships in the rest of the treatments, the
 226 subset of COG and activity-abundance-nutrient variables from control, DOMhum, DOMagri, NP and LowO₂
 227 were analyzed independently (Fig. 5, Table S4). In Exp I three independent clusters were observed (Fig. 5A).
 228 The first cluster of COGs was associated with protease activity (co-related with alkaline phosphatase activity
 229 and PO₄ concentration) and characterized by COGs belonging to 'Carbohydrate transport and metabolism'
 230 and 'General function' categories. The second cluster of COGs related to amino acid uptake (co-related to

DOC and ammonium concentration) and was enriched with 'Mobilome: prophages, transposons', primarily phage related proteins (Table S4). Other strongly associated COGs were involved in pyrimidine metabolism and amino acid transport. The third cluster of COGs was strongly associated to BP, containing mostly COG associated with unknown or general functions, except included three ABC transporters and COGs related to carbohydrate metabolism pathways (Table S4).

In Exp II, COG clusters and associated variables were distinct from the association analysis of Exp I (Fig. 5B, Table S4). One cluster of COGs associated with alkaline phosphatase activity and was enriched with COGs belonging to the 'General function' category including two ABC-type transporters. The COGs associated with protease activity belonged to 'Carbohydrate transport and metabolism' and 'Function unknown', and included two proteases. The third cluster of COGs which was strongly associated with amino acid uptake, BA and BP contained COGs belonging to the category 'Inorganic ion transport and metabolism'. Other strongly related COGs in this cluster were enzymes involved in purine degradation and in nitrate assimilation. β -glucosidase activity was significantly related to DCA gradients ($r^2 = 0.69$, $p = 0.001$) but was not associated with any COG clusters.

Discussion

Prokaryotic communities are capable of adjusting to the physical and chemical factors defining a specific environment. Consequently, functional differences in microbial communities can be used to discriminate qualitatively dissimilar environments (Tringe *et al.*, 2005; Dinsdale *et al.*, 2008; Gianoulis *et al.*, 2009). Indeed, our results document selection for unique functional metagenomic patterns in marine microbial communities in response to selected environmental stressors at two geographic sites. We were further able to link the changes in metagenomic profiles to dynamics in microbial activities caused by environmental perturbations. This shows how adaptations to environmental changes can shape both microbial community structure and functional metagenomic profiles over a short time scale.

Site specific blueprint

257 The major differences in key metabolic pathways at the two sampling sites was in agreement with previous
258 observations of elevated abundance of ED and EMP pathway genes in high and low salinity environments,
259 respectively (Dupont *et al.*, 2014; Fortunato and Crump, 2015). This was further supported by the high
260 abundances of *Gammaproteobacteria* in Exp. I (high salinity) and high abundances of *Actinobacteria* in Exp
261 II (low salinity), as these bacterial groups previously have been found to be associated with the ED and EMP
262 pathway, respectively (Flamholz *et al.*, 2013; Fortunato and Crump, 2015). The observed relationship
263 between salinity, microbial community composition and metabolic function suggested that salinity was an
264 important environmental factor shaping the metagenomic blueprint at the two sites, as also previously shown
265 for microbial communities in the BS (Herlemann *et al.*, 2011; Dupont *et al.*, 2014).
266 In general, the number and composition of COGs that changed due to manipulations differed between the
267 two sites suggesting that the distinct communities showed different adaptations to the same environmental
268 stressors. However, the distinct responses may also reflect differences in other environmental parameters
269 such as temperature and seasonal variations in the composition of autochthonous and allochthonous DOM
270 added to the experiments.

271 272 *Changes in metagenomes in response to treatments*

273 The significant changes in the metagenomes associated with the applied environmental stressors support
274 previous suggestions that different sources of bioavailable DOC induce different genomic and functional
275 responses in a microbial community (McCarren *et al.*, 2010; Poretsky *et al.*, 2010). Our study indicates
276 several links between changes in abundance of specific groups of functional genes and differences in
277 community activity and environmental parameters. For example, the elevated C:N ratio resulting from DOM
278 input from the humic river in Exp I was associated with an increase in COGs related to carbohydrate
279 metabolism, indicating the processing of new sources of bioavailable DOC. Furthermore, increased
280 abundance of COGs involved in phosphate transport suggested that the input of humic matter stimulated
281 phosphate uptake as previously suggested (Carlsson *et al.*, 1993). The shared COGs in the DOMhum

282 treatments belonged mainly to TRAP transporters and carbohydrate metabolism pathways, demonstrating
283 that the added humic DOM increased the abundance of COGs involved in specific DOC assimilation and
284 consumption pathways in both experiments. TRAP-type transporters transfer low molecular weight
285 extracellular solutes (Forward *et al.*, 1997), especially organic acids (Mulligan *et al.*, 2011), into the cell.
286 These compounds are likely products of extracellular hydrolysis of high molecular weight humic material,
287 and may therefore be indicators of the input of humic DOM with river water.

288 Interestingly, we observed that DOMagri and NP treatments both resulted in up-representation of COGs that
289 correlated with amino acid uptake, NH_4 and DOC concentration (Table S2 and S4), suggesting that the
290 increased nitrate and ammonium loads shaped similar genetic responses related to N-metabolism in these
291 treatments. In Exp II, on the other hand, similar change in metagenome was observed in DOMagri and
292 DOMhum, characterized by a stimulation of DOC transporters (e.g. particularly TRAP and ABC
293 transporters). This suggested that the bacterial community was mainly responding to the DOC input from the
294 river water. The simultaneous net uptake of inorganic N in these treatments corresponded with high C:N
295 ratio and N limitation and was reflected in increased abundance of COGs involved in nitrogen fixation.

296 Increased prevalence of COGs related to proteases in the DOMsea treatments (i.e. the top three most up-
297 represented COGs) was in agreement with the increased protease activity, suggesting that the DOM from the
298 natural plankton lysate stimulated protein metabolism. Natural plankton lysate represents a variety of organic
299 compounds that is readily recycled by heterotrophs (Amon and Benner, 1996) as indicated by the elevated
300 protease activity. Likewise, LTC treatments showed the largest effect on bacterial metagenomic blueprints in
301 both experiments, with dominance of COGs associated with general metabolic functions. This was also
302 reflected in the highly elevated microbial activity and abundance associated with the significant increase in
303 bioavailable DOM. Cells lysates are rich in proteins and DNA and can constitute a high quality substrate
304 source for heterotrophic bacteria (Middelboe and Jørgensen, 2006; Haaber and Middelboe, 2009), as also
305 indicated by the elevated protease, glucosidase and phosphatase activity in these treatments. This change of
306 metabolic properties was also associated with shifts in microbial community composition in both LTC
307 experiments towards dominance of opportunistic and fast-growing *Gammaproteobacteria* (Eilers *et al.*,

2000; Pinhassi and Berman, 2003) known for their ability to exploit DOM when available (Fuchs *et al.*, 2000; Herlemann *et al.*, 2014).

The LowO₂ treatment showed generally few metagenomic changes relative to the control in both experiments, in agreement with the similar development in bacterial abundance and activities (e.g. BA, BP, and enzymatic activities) and low or unchanged DOC consumption measured for this treatment compared to control. Consequently, within the experimental time frame, the substrate concentration (DOM lability) and composition (e.g. C:N:P ratio, humic components and protein content) had a much more pronounced effect on the bacterial community than reducing oxygen concentration to 15% saturation.

Some of the most up-represented COGs in the treatments compared to controls involved regulatory mechanisms such as transcriptional regulators. These were especially observed in the DOM treatments, e.g. DOMhum, DOMagri and LTC, indicating an important role of these COGs for rapid adaptations to changing substrate conditions (Cases *et al.*, 2003). Likewise, several phage related proteins and transposases were up-represented in these treatments compared to controls. The function of these COGs may be associated with horizontal gene transfer, a common source of microbial diversity and genomic plasticity, which is often enriched in highly dynamic environments (Angly *et al.*, 2009). Alternatively, the up-representation may reflect increased induction of prophages following exposure to changing environmental conditions or by stimulation of bacterial metabolic activity.

Effects of environmental changes on bacterial transporters

In order to successfully attain essential nutrients and DOM from marine environments with limited resources, bacteria need efficient transport systems. Bacterial transporters are frequently observed in microbial metatranscriptome and metaproteome data (McCarren *et al.*, 2010; Morris *et al.*, 2010; Poretsky *et al.*, 2010) and have been predicted to account for between 3% and 16% of the genes in prokaryote genomes (Angly *et al.*, 2009). Therefore, we analyzed abundance change of various transporter types in the metagenomes, grouping the most abundant transporters annotated using COGs (Fig. 8). The large increase of transporter abundance associated with LTC, DOMhum and DOMsea treatments in Exp I indicated that

transporters were particularly responsive to input of organic matter rather than inorganic nutrients. In Exp II on the other hand, DOM additions stimulated a variety of different transporters that in some cases were specific for the specific DOM sources, e.g. an increased abundance of ABC-transporters associated with uptake of complex carbohydrates and N-compounds in the DOMhum treatment (Fig. 4). This corresponds well with previous indications that certain ABC-type transporters are responsible for uptake of possibly refractory DOM in the BS (Jiao and Zheng, 2011) and such transporters may therefore be useful as indicators of increased load of refractory material. DOMsea treatments from both Exp I and II showed up-represented TonB-dependent transporters, which are involved in bacterial uptake of scarce resources such as iron-complexes (e.g. siderophores, heme, etc.), vitamin B12, and some carbohydrates from nutrient-limiting environments (Tang *et al.*, 2012). These transporters are often associated with Flavobacteria (Teeling, 2012), which were also dominant in our experiments (data not shown).

345

346 *Linking activity with genetic blueprints*

347 We found significant correlations between up-represented COGs and the microbial and chemical parameters
348 in the DOMhum, DOMagri, DOMsea, LowO₂ and NP treatments (Fig. 5, Table S4) allowing us to associate
349 the change in metagenomes to biochemical processes at various environmental conditions. Most prominently
350 alkaline phosphatase and protease activity correlated with several COGs involved in metabolic pathways
351 responsible for transport and metabolism of inorganic compounds, carbohydrates, and organic N-compounds
352 (i.e. amino acids). Amino acid uptake correlated with several COGs involved in nitrogen metabolism
353 pathways, in particular purine degradation and pyrimidine biosynthesis. These COGs were found mainly in
354 DOMhum and DOMagri treatments suggesting higher availability of purines and pyrimidines in the river
355 water.

356 In conclusion, we demonstrated that bacterioplankton communities responded significantly to the
357 environmental disturbances in terms of activity and genetic functional characteristics humic. Notably, the
358 metagenomic and metabolic changes in bacterial communities could be ranked in response to various
359 stressors. The most prominent response occurred to changes in DOM load from river water, and depended on

the catchment area, i.e. humic and agricultural influenced catchments. Cyanobacterial lysates affected the microbial activity and metagenome mainly via an increased load of labile organic compounds leading to increased turnover of organic compounds and remineralization. Moderate decrease in oxygen concentration did not affect the metabolism or genetic composition of bacterial communities within the relatively short time frame of the experiments. The complex and diverse effects of the experimental manipulations on the bacterial metagenome and metabolic activity highlight the challenge of linking specific metagenomic changes to specific environmental conditions in the BS. Nonetheless, our experiments allowed us to identify key characteristics of metagenomic profiles in response to environmental changes and link these to process rates. This provides a first step in exploring the use of bacterial gene abundance as indicators of biogeochemical processes and environmental conditions and advances our exploration of the potential use of marine microorganisms as markers of monitoring environmental conditions.

371

372 **Experimental procedures**

373 *Study sites*

374 Seawater (300 L) was collected from 5 m depth in Øresund (56°3'26" N, 12°38'45" E; Exp I) on 20th of April
375 2015 and Storfjärden (Gulf of Finland; 59°51'12" N 23°16'19" E; Exp II) 27th of July 2015. Both sites are
376 characterized by low salinity (13.4 and 6.0 for Øresund and Storfjärden, respectively) and *in situ* temperature
377 was 7°C in April and 14°C in July (Table 1). River water was collected 2 days before the experiments from
378 the humic influenced Lapväärtti river (62°14'21" N 21°34'38" E, Finland) and the Lielupe river (56°48'42" N
379 23°35'5" E, Latvia) which is mainly influenced by agricultural runoff (Fig. S1). River water was filtered
380 through 0.22 µm capsule filters (Optical XL filter, Millipore) and kept in the dark at 4°C until use. The
381 salinity of the river water was adjusted to Øresund or Storfjärden *in situ* concentrations with muffled NaCl.

382 *Experimental design*

383 Seawater was transported in acid-washed Milli-Q rinsed polycarbonate bottles to a climate room with *in situ*
384 temperature within 1 h. The seawater was filtered in two ways (Fig. S8): (i) to prepare inoculum for the batch
385 incubations, water was sieved through a 10 µm plankton net; (ii) for generating bacteria free water, which

386 was obtained by additionally filtering water through 0.22 μ m capsule filters (Optical XL filter, Millipore),
 387 and subsequently served as medium for the treatments. Recovered material collected on the 10 μ m plankton
 388 net during filtration step (i) (from 60 L) was collected and resuspended in 50 ml 0.22 μ m filtered seawater,
 389 sonicated and sterile filtered with a 0.22 μ m Sterivex capsule filter (Millipore) and used as autochthonous
 390 DOM source for DOMsea treatments. Cyanobacterial lysate was prepared by freezing and thawing cultivated
 391 *Nodularia spumigena* UHCC 0039 cells (University of Helsinki Culture Collection). Lysate was 0.22 μ m
 392 filtered prior to use.

393 The seven treatments were prepared and inoculated as followed (Fig. S8): 1) DOMhum; addition of 20%
 394 0.22 μ m filtered humic rich river water from Lapväärtti, 2) DOMagri; addition of 20% 0.22 μ m filtered
 395 agricultural river water from Lielupe, 3) DOMsea; addition of autochthonous DOM source, 4) LTC; addition
 396 of lysate from toxic cyanobacteria, 5) NP; addition of inorganic nutrients to final concentration of 1 μ M
 397 phosphate, 1 μ M ammonia and 9 μ M nitrate, 6) LowO₂; oxygen concentration reduced to ~15% saturation
 398 by bubbling with N₂-gas, and 7) control; no manipulation. The X L microcosms (six treatments in triplicates
 399 and controls in six replicates) containing bacteria-free water were inoculated with bacteria at a ratio of 1:5
 400 and incubated in the dark in a climate room at *in situ* temperature for 3-4 days. Mixing was provided by
 401 continuous bubbling with atmospheric air (with the exception of LowO₂ treatments). *In situ* collected
 402 samples and the manipulated batch treatments were initially sub-sampled for measurements of bacterial
 403 abundance (BA) and production (BP), O₂ consumption, inorganic nutrient and organic matter concentrations
 404 (DOC/DON), amino acid uptake and activity of bacterial extracellular enzymes at time zero, and
 405 subsequently with 6 to 12 h intervals in the batch experiments. For analysis of bacterial community
 406 composition, samples were collected from the last time point (T_{end}, 78 h for Exp I and 96 h for Exp II) on
 407 0.22 μ m filters for 16S rRNA gene fragment sequencing and for total community DNA based metagenomics.
 408 Full details on bacterial and viral abundance (VA), bacterial production (BP), amino acid uptake, DOC/DON
 409 concentrations, oxygen saturation, extracellular enzymes activity and chemical analyses are provided in
 410 *Experimental procedures S1*.
 411 *Nucleic acids extraction*

Water samples for DNA/RNA were collected in 1L PC bottles (Nalgene) and fixed in a 10% final volume stop-solution (5% of phenol in 99.8% ethanol). The bottles were inverted to ensure complete mixing and stored under dark conditions until filtration (max 24 h). Samples were filtered onto 0.22 µm Durapore GVWP04700 filters (Millipore), which were stored at -80°C. Prior to DNA/RNA extractions, the filters were cut and vortexed for 10 min with UV treated zirconia beads and RLT Plus Buffer containing β-mercaptoethanol. Total DNA/RNA was isolated using the AllPrep DNA/RNA Mini Kit (Qiagen) and on-column DNA digestion (RNase-Free DNase Set, Qiagen) was applied. The second DNA digestion was carried out using the TURBO DNA-free™ kit (Life Technologies) following manufacturer's instructions. The DNA Clean & Concentrator™ and RNA Clean & Concentrator™ kits (Zymo Research) were used to purify and concentrate total genomic DNA. DNA concentrations were measured using Quant-IT PicoGreen (Invitrogen) and Quant-IT RiboGreen (Invitrogen) assays, respectively.

Sequencing and metagenomic analysis

DNA (2–10 ng) from each sample was prepared with the Rubicon ThruPlex kit (Rubicon Genomics, Ann Arbor, Michigan, USA) according to the instructions of the manufacturer. Cleaning steps were performed with MyOne™ carboxylic acid-coated superparamagnetic beads (Invitrogen, Carlsbad, CA, USA). Libraries were sequenced on a HiSeq 2500 (Illumina Inc., San Diego, CA, USA). On average, 18 million paired-end reads of 2 × 125 bp per sample were generated. Raw reads were quality trimmed with Cutadapt (Martin, 2011) from both read ends and duplicate reads were removed with fastuniq (Xu *et al.*, 2012). High quality reads were mapped onto the BARM database containing the most informative reference genomes in the BS (Alneberg *et al.*, submitted) with Bowtie2 using default parameters (Langmead *et al.*, 2013). The raw counts were calculated from Bedtools histogram output (Quinlan, 2014) and quantitative abundance of reads were annotated using COG protein families (Galperin *et al.*, 2015).

For determining the community composition partial 16S rRNA genes (V3- V5 region) were PCR amplified using the primers Bakt_341F (CCTACGGGNGGCWGCA) and Bakt_805R (GACTACHVGGGTATCTAATCC) (Herlemann *et al.*, 2011). Products were then sequenced with an Illumina MiSeq paired-end multiplex platform in SciLifeLab/NGI (Solna, Sweden).

438 Raw amplicon reads were quality trimmed with Trimmomatic (ver 0.32). We removed any chimeric reads
439 and assigned the reads into operational taxonomic units (OTUs) using 97% cut-off in cd-hit-otu (Li *et al.*,
440 2012). We used SINA 1.2 (Pruesse *et al.*, 2012) against the SILVA database (v. 115) to classify unique
441 OTUs and estimated the relative abundance of each OTU using an in-house Python script.

442 443 *Data analyses*

444 All data and statistical analyses were carried out in R (<http://www.r-project.org/>). The EdgeR package was
445 used for differential abundance analysis (Robinson *et al.*, 2010) since the data followed a negative binomial
446 distribution. Low abundance COGs were filtered out by only including COGs exceeding 100 counts per
447 million in at least three samples, leaving a total of 2,065 and 2,015 COGs for further analysis for Exp I and
448 Exp II, respectively. Trimmed mean of M-values (TMM) normalization was performed to eliminate
449 composition biases between libraries. The exact test function was used in order to identify COGs that
450 displayed significantly difference in normalized abundance level between control replicates and treatment
451 triplicates. We extracted differentially abundant COGs at a significance level of $p < 0.01$ and with a false
452 discovery rate (FDR) of < 0.05 . As an additional quality control step of COGs detected with significantly
453 different abundance for the cyanobacterial lysate treatment, we required that the logFC had to be more than
454 1.5 compared to the average of controls. A high number of significant COGs for treatments from Exp II
455 (DOMhum, DOMagri, DOMsea and NP) were identified with these criteria. To shortlist these, only COGs
456 with logFC more than 0.4 compared to the average of controls was extracted.

457 Cluster analyses were based on normalized abundance profiles of significantly up-represented COGs
458 identified as described above ($n = 221$ for Exp I and $n = 144$ for Exp II) and performed using Pearson
459 correlation and hclust function in R. For k-medoids clustering, partitioning around medoid (PAM) algorithm
460 was used with Pearson correlation for distance calculations. The number of clusters was validated using
461 silhouettes (data not shown) (Rousseeuw, 1987).

462 Transporter examination was performed by extracting transporters from the subset of significantly up
463 represented COGs ($\logFC > 0$, $FDR < 0.05$, $p\text{-adjusted} < 0.001$) based on previous indications of their

464 relevance to biogeochemical processes (Poretsky *et al.*, 2010; Gifford *et al.*, 2013; Harke and Gobler, 2013;
465 Smith *et al.*, 2013; Penn *et al.*, 2014; Saito *et al.*, 2014; Satinsky *et al.*, 2014; Shilova *et al.*, 2014; Doxey *et*
466 *al.*, 2015).

467 Linear mixed models were fitted to the time series measurements of BA, VA, BP, enzyme activities, amino
468 acid uptake, DOC/DON and inorganic nutrient concentrations using the R package lme4 (Bates *et al.*, 2015).
469 Levels of variables were modelled using the treatment and time as fixed effects and replicates of microcosms
470 as random effects. Significance of fixed effects was assessed by an F-test using a significance level of 5%.
471 Model checking was based on residual plots and normal probability plots using the raw residuals. Models
472 were reduced using the likelihood ratio test. A 1% or 5% significance level was used. Pairwise comparisons
473 were evaluated based on adjusted *p*-values obtained using the single-step method comparing control
474 treatment with treatments (Table S1). Detrended correspondence analysis (DCA) of COG abundance data
475 and linear fitting of measured abundance/rate/concentration variables was analyzed using vegan package in
476 R. Diversity indices such as abundance-based coverage estimator (ACE) (Chao and Lee, 1992) and Shannon
477 index (Hill, 1973) were calculated and differences between treatments were tested statistically using
478 ANOVA.

479

480 **Acknowledgements**

481 The authors would like to thank J. Hansen for technical assistance, S. J. Traving for useful discussions, C.
482 Ritz for statistical advice, and the staff of R/V Ophelia and Tvärminne Zoological Station for technical
483 assistance. This work resulted from the BONUS Blueprint project supported by BONUS (Art 185), funded
484 jointly by the EU and the Danish Council for Independent Research, Estonian Research Council, , Swedish
485 Research Council FORMAS, and Academy of Finland. In addition V.K. was supported by a personal grant
486 PUT-134 and 1389 from the Estonian Research Council.

487 Table 1. *In situ* abundance and activity of bacterioplankton and viruses, concentrations of nutrients and
 488 environmental conditions. Nd = no data

	Øresund (Exp I)	Storfjärden (Exp II)	Lapväärtti river	Lielupe River
Sampling time	April	August	April	April
Bacterial abundance ($\times 10^6$ cells mL ⁻¹)	2.0 $\pm$ 2.5	3.3 $\pm$ 2.6	nd	nd
Bacterial production (μ g C L ⁻¹ d ⁻¹)	1.95 $\pm$ 0.06	4.1 $\pm$ 0.3	nd	nd
Viral abundance ($\times 10^6$ viruses mL ⁻¹)	3.2 $\pm$ 4.9	15.5 $\pm$ 2.0	nd	nd
DOC (μ M)	381 $\pm$ 7	667 $\pm$ 91	1561 $\pm$ 16	1337 $\pm$ 30
DON (μ M)	50 $\pm$ 2	50 $\pm$ 2	78.0 $\pm$ 0.3	429 $\pm$ 16
NH ₄ ⁺ (μ M)	0.24 $\pm$ 0.01	0.26 $\pm$ 0.01	1.82 $\pm$ 0.02	1.77 $\pm$ 0.06
NO ₃ ⁻ (μ M)	2.1 $\pm$ 0.2	0.44 $\pm$ 0.03	22.2 $\pm$ 0.5	548 $\pm$ 47
PO ₄ ³⁻ (μ M)	0.09 $\pm$ 0.02	0.26 $\pm$ 0.03	0.37 $\pm$ 0.04	0.30 $\pm$ 0.07
Salinity	13.4	6.0	0	0
Temperature (°C)	7	14	nd	nd

489

Table 2. Identification of COGs up- or down-represented in response to treatment. DOMhum, fresh water from humic rich river; DOMagri, fresh water from polluted river influenced by agriculture; DOMsea autochthonous dissolved organic matter; LTC, cell lysate from cyanobacterial culture; NP inorganic nutrient addition; LowO₂, reduced O₂ concentration. Extraction of significant COGs was performed with different filtering criteria; ^a Criteria for extracted significant COGs were $p < 0.01$, FDR < 0.05 . ^b Criteria for extracted significant COGs were $p < 0.01$, FDR < 0.05 , logFC > 1.5 . ^c Criteria for extracted significant COGs were $p < 0.01$, FDR < 0.05 , logFC > 0.4 . ^d Average of normalized counts (triplicates) for all significant COGs ($p < 0.01$, FDR < 0.05). NF indicates that no additional filter was applied on treatment.

	Treatments	Number of significant COGs ^a (% of all)	Total counts of significant COGs (cpm) ^d	Number of significant COGs after filtering	Up-represented COGs	Down-represented COGs	%DOC-related COGs of up-represented
Exp I	DOMhum	207 (10%)	120226	NF	132	75	36%
	DOMagri	16 (1%)	6380	NF	14	2	21%
	DOMsea	23 (1%)	18382	NF	20	3	65%
	LTC	1480 (72%)	762058	167 (8%) ^b	57	108	37%
	NP	15 (1%)	6919	NF	14	1	21%
	LowO ₂	40 (2%)	7912	NF	22	18	18%
Exp II	DOMhum	361 (18%)	188767	83 (4%) ^c	56	27	46%
	DOMagri	368 (18%)	151391	166 (8%) ^c	43	123	40%
	DOMsea	268 (13%)	132625	90 (4%) ^c	29	61	52%
	LTC	1525 (76%)	795492	145 (7%) ^b	47	98	40%
	NP	390 (19%)	173299	95 (4%) ^c	21	74	33%
	LowO ₂	13 (1%)	6843	NF	10	3	30%

Figure legends

Fig. 1. Time course of bacterial abundance (BA, upper panel), viral particles abundance (VA, middle panel) and bacterial production (BP, bottom panel) in treatments for Exp I (left panel) and Exp II (right panel). DOMhum, fresh water from humic rich river; DOMagri, fresh water from polluted river influenced by agriculture; DOMsea, autochthonous dissolved organic matter; LTC, cell lysate from cyanobacterial culture; NP, inorganic nutrient addition; LowO₂, reduced O₂ concentration; Control, controls. Significant difference between treatments and controls were tested using Dunnett's test (Table S1).

Fig. 2. Extracellular enzymatic activities in microcosms from Exp I (left panel) and Exp II (right panel). Protease activity are shown in upper panel), β -glucosidase activity in middle panel and alkaline phosphatase activity in the bottom panel. DOMhum, fresh water from humic rich river; DOMagri, fresh water from polluted river influenced by agriculture; DOMsea, autochthonous dissolved organic matter; LTC, cell lysate from cyanobacterial culture; NP, inorganic nutrient addition; LowO₂, reduced O₂ concentration; Control, controls. Significant difference between treatments and controls were tested using Dunnett's test (Table S1).

Fig. 3. Microbial community composition assessed by 16S rRNA gene sequence analysis of treatments from Exp I (A) and treatments from Exp II (B) at time point T_{end} (T₇₈ for Exp I and T₉₆ for Exp II). Clustering of COGs significantly up represented in one or more treatments compared to controls for (C) Exp I (n = 221) and (D) Exp II (n = 144). Heatmap coloring reflects the Z score of the normalized abundances (cpm) of each COG across all clustered data. The same COGs have been clustered according to treatment into five clusters by k-medoids clustering for Exp I (E) and Exp II (F). The y axis displays COG abundances and the x axis shows the different treatments. DOMhum, fresh water from humic rich river; DOMagri, fresh water from polluted river influenced by agriculture; DOMsea, autochthonous dissolved organic matter; LTC, cell lysate from cyanobacterial culture; NP, inorganic nutrient addition; LowO₂, reduced O₂ concentration; Control, controls.

Fig. 4. Increase (as logFC) in abundance of significantly changing transporter COGs compare to controls classified into transporter type (A), and compounds involved in activity of these transporters (B). DOMhum, fresh water from humic rich river; DOMagri, fresh water from polluted river influenced by agriculture; DOMsea, autochthonous dissolved organic matter; LTC, cell lysate from cyanobacterial culture; NP, inorganic nutrient addition; LowO₂, reduced O₂ concentration.

Fig. 5. Association based on detrended correspondence analysis between relative abundance of reads in COGs and measured activities, rates and concentrations of nutrients in: (A) Exp I: BP ($r^2 = 0.2518$, $p = 0.088$), DOC ($r^2 = 0.42$, $p = 0.007$), Ammonium ($r^2 = 0.38$, $p = 0.019$), PO₄ ($r^2 = 0.37$, $p = 0.095$), protease activity ($r^2 = 0.50$, $p = 0.004$), alkaline phosphatase ($r^2 = 0.34$, $p = 0.026$). BA, VA, DON, NO₃ and β -glucosidase were not significant. (B) Exp II: BP ($r^2 = 0.85$, $p = 0.001$), BA ($r^2 = 0.62$, $p = 0.002$), VA ($r^2 = 0.33$, $p = 0.027$), DOC ($r^2 = 0.26$, $p = 0.086$), Protease activity ($r^2 = 0.79$, $p = 0.001$), β -glucosidase ($r^2 = 0.69$, $p = 0.001$), alkaline phosphatase ($r^2 = 0.42$, $p = 0.010$). DON, Ammonium, PO₄ and NO₃ were not significant. Pie charts indicate proportions of COG clustered into COG categories which are related with measured activities, rates and concentrations of nutrients.

Reference

- Amon, R.M.W. and Benner, R. (1996) Bacterial utilization of different size classes of dissolved organic matter. *Limnol. Oceanogr.* **41**: 41–51.
- Anantharaman, K., Brown, C.T., Hug, L.A., Sharon, I., Castelle, C.J., Probst, A.J., et al. (2016) Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* **7**: 13219.
- Andersson, A.F., Riemann, L., and Bertilsson, S. (2009) Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. *ISME J.* **4**: 171–181.
- Angly, F.E., Willner, D., Prieto-Davó, A., Edwards, R.A., Schmieder, R., Vega-Thurber, R., et al. (2009) The GAAS metagenomic tool and its estimations of viral and microbial average genome size in four major biomes. *PLoS Comput. Biol.* **5**: e1000593.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015) Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**: 51.
- Burke, C., Steinberg, P., Rusch, D., Kjelleberg, S., and Thomas, T. (2011) Bacterial community assembly based on functional genes rather than species. *PNAS* **108**: 14288–14293.
- Caruso, G., La Ferla, R., Azzaro, M., Zoppini, A., Marino, G., Petochi, T., et al. (2016) Microbial assemblages for environmental quality assessment: Knowledge, gaps and usefulness in the European Marine Strategy Framework Directive. *Crit. Rev. Microbiol.* **42**: 883–904.
- Cases, I., De Lorenzo, V., and Ouzounis, C.A. (2003) Transcription regulation and environmental adaptation in bacteria. *Trends Microbiol.* **11**: 248–253.
- Chao, A. and Lee, S.-M. (1992) Estimating the Number of Classes via Sample Coverage. *J. Am. Stat. Assoc.* **87**: 210–217.
- Dinsdale, E. a, Edwards, R. a, Hall, D., Angly, F., Breitbart, M., Brulc, J.M., et al. (2008) Functional metagenomic profiling of nine biomes. *Nature* **452**: 629–632.
- Doxey, A.C., Kurtz, D.A., Lynch, M.D., Sauder, L.A., and Neufeld, J.D. (2015) Aquatic metagenomes implicate Thaumarchaeota in global cobalamin production. *ISME J* **9**: 461–471.
- Dupont, C.L., Larsson, J., Yooseph, S., Ininbergs, K., Goll, J., Asplund-Samuelsson, J., et al. (2014) Functional tradeoffs underpin salinity-driven divergence in microbial community composition. *PLoS One* **9**: e0089549.
- Eilers, H., Pernthaler, J., and Amann, R. (2000) Succession of pelagic marine bacteria during enrichment: A close look at cultivation-induced shifts. *Appl. Environ. Microbiol.* **66**: 4634–4640.
- Farnelid, H., Bentzon-Tilia, M., Andersson, A.F., Bertilsson, S., Jost, G., Labrenz, M., et al. (2013) Active nitrogen-fixing heterotrophic bacteria at and below the chemocline of the central Baltic Sea. *ISME J.* **7**: 1413–23.
- Flamholz, A., Noor, E., Bar-Even, A., Liebermeister, W., and Milo, R. (2013) Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc. Natl. Acad. Sci. U. S. A.* **110**: 10039–44.
- Fortunato, C.S. and Crump, B.C. (2015) Microbial gene abundance and expression patterns across a river to ocean salinity gradient. *PLoS One* **10**: e0140578.
- Forward, J.A., Behrendt, M.C., Wyborn, N.R., Cross, R., and Kelly, D.J. (1997) TRAP transporters: A new family of periplasmic solute transport systems encoded by the dctPQM genes of *Rhodobacter capsulatus* and by homologs in diverse gram-negative bacteria. *J. Bacteriol.* **179**: 5482–5493.
- Franzosa, E. a., Hsu, T., Sirota-Madi, A., Shafquat, A., Abu-Ali, G., Morgan, X.C., and Huttenhower, C. (2015) Sequencing and beyond: integrating molecular “omics” for microbial community profiling. *Nat. Rev. Microbiol.* **13**: 360–72.

586 Fuhrman, J.A., Cram, J.A., and Needham, D.M. (2015) Marine microbial community dynamics and their
587 ecological interpretation. *Nat. Rev. Microbiol.* **13**: 133–146.

588 Galperin, M.Y., Makarova, K.S., Wolf, Y.I., and Koonin, E. V. (2015) Expanded Microbial genome
589 coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* **43**: D261–
590 D269.

591 Gianoulis, T.A., Raes, J., Patel, P. V, Bjornson, R., Korbel, J.O., Letunic, I., et al. (2009) Quantifying
592 environmental adaptation of metabolic pathways in metagenomics. *Proc. Natl. Acad. Sci.* **106**: 1374–
593 1379.

594 Gifford, S.M., Sharma, S., Booth, M., and Moran, M.A. (2013) Expression patterns reveal niche
595 diversification in a marine microbial assemblage. *ISME J.* **7**: 281–298.

596 Gilbert, J. a and Dupont, C.L. (2011) Microbial metagenomics: beyond the genome. *Ann. Rev. Mar. Sci.* **3**:
597 347–371.

598 Gomez-Consarnau, L., Lindh, M. V, Gasol, J.M., and Pinhassi, J. (2012) Structuring of bacterioplankton
599 communities by specific dissolved organic carbon compounds. *Env. Microbiol* **14**: 2361–2378.

600 Graham, E.B., Knelman, J.E., Schindlbacher, A., Siciliano, S., Breulmann, M., Yannarell, A., et al. (2016)
601 Microbes as engines of ecosystem function: When does community structure enhance predictions of
602 ecosystem processes? *Front. Microbiol.* **7**: 214. doi: 10.3389/fmicb.2016.00214.

603 Grote, J., Schott, T., Bruckner, C.G., Glöckner, F.O., Jost, G., Teeling, H., et al. (2011) Genome and
604 physiology of a model Epsilonproteobacterium responsible for sulfide detoxification in marine oxygen
605 depletion zones. *Proc. Natl. Acad. Sci. U. S. A.* **109**: 506–10.

606 Haft, D.H., Selengut, J.D., Richter, R.A., Harkins, D., Basu, M.K., and Beck, E. (2013) TIGRFAMs and
607 genome properties in 2013. *Nucleic Acids Res.* **41**: D387–D395. doi: 10.1093/nar/gks1234.

608 Harke, M.J. and Gobler, C.J. (2013) Global Transcriptional Responses of the Toxic Cyanobacterium,
609 *Microcystis aeruginosa*, to Nitrogen Stress, Phosphorus Stress, and Growth on Organic Matter. *PLoS*
610 *One* **8**: e0069834.

611 Herlemann, D.P., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J.J., and Andersson, A.F. (2011)
612 Transitions in bacterial communities along the 2000km salinity gradient of the Baltic Sea. *ISME J.* **5**:
613 1571–1579.

614 Herlemann, D.P.R., Lundin, D., Labrenz, M., Jürgens, K., Zheng, Z., Aspeborg, H., and Andersson, A.F.
615 (2013) Metagenomic De Novo assembly of an aquatic representative of the verrucomicrobial class
616 *Spartobacteria*. *MBio* **4**: e00569-12.

617 Hill, M.O. (1973) Diversity and evenness: a unifying notation and its consequences. *Ecology* **54**: 427–432.

618 Haaber, J. and Middelboe, M. (2009) Viral lysis of *Phaeocystis pouchetii*: implications for algal population
619 dynamics and heterotrophic C, N and P cycling. *Isme J* **3**: 430–441.

620 Jiao, N. and Zheng, Q. (2011) The microbial carbon pump: From genes to ecosystems. *Appl. Environ.*
621 *Microbiol.* **77**: 7439–7444.

622 Kirchman, D.L., Dittel, A.I., Findlay, S.E.G., and Fischer, D. (2004) Changes in bacterial activity and
623 community structure in response to dissolved organic matter in the Hudson River, New York. *Aquat.*
624 *Microb. Ecol.* **35**: 243–257.

625 Kisand, V., Valente, A., Lahm, A., Tanet, G., and Lettieri, T. (2012) Phylogenetic and functional
626 metagenomic profiling for assessing microbial biodiversity in environmental monitoring. *PLoS One* **7**:
627 e0043630.

628 Labrenz, M., Sintes, E., Toetzke, F., Zumsteg, A., Herndl, G.J., Seidler, M., and Jürgens, K. (2010)
629 Relevance of a crenarchaeotal subcluster related to *Candidatus Nitrosopumilus maritimus* to ammonia
630 oxidation in the suboxic zone of the central Baltic Sea. *ISME J.* **4**: 1496–1508.

631 Landa, M., Cottrell, M.T., Kirchman, D.L., Kaiser, K., Medeiros, P.M., Tremblay, L., et al. (2014)
632 Phylogenetic and structural response of heterotrophic bacteria to dissolved organic matter of different
633 chemical composition in a continuous culture study. *Environ. Microbiol.* **16**: 1668–1681.

634 Langmead, B., Salzberg, S.L., and Langmead (2013) Bowtie2. *Nat. Methods* **9**: 357–359.

635 Lauro, F.M., McDougald, D., Thomas, T., Williams, T.J., Egan, S., Rice, S., et al. (2009) The genomic basis
636 of trophic strategy in marine bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **106**: 15527–33.

637 Li, W., Fu, L., Niu, B., Wu, S., and Wooley, J. (2012) Ultrafast clustering algorithms for metagenomic
638 sequence analysis. *Brief. Bioinform.* **13**: 656–668.

639 Lindh, M. V., Figueroa, D., Sjöstedt, J., Baltar, F., Lundin, D., Andersson, A., et al. (2015) Transplant
640 experiments uncover Baltic Sea basin-specific responses in bacterioplankton community composition
641 and metabolic activities. *Front. Microbiol.* **6**: 223. doi: 10.3389/fmicb.2015.00223.

642 Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads.
643 *EMBnet.journal* **17**: 10.

644 McCarren, J., Becker, J.W., Repeta, D.J., Shi, Y., Young, C.R., Malmstrom, R.R., et al. (2010) Microbial
645 community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic
646 matter turnover in the sea. *Proc. Natl. Acad. Sci. U. S. A.* **107**: 16420–7.

647 Middelboe, M. and Jørgensen, N.O.G. (2006) Viral lysis of bacteria: an important source of dissolved amino
648 acids and cell wall compounds. *J. Mar. Biol. Assoc. UK* **86**: 605–612.

649 Morris, R.M., Nunn, B.L., Frazar, C., Goodlett, D.R., Ting, Y.S., and Rocap, G. (2010) Comparative
650 metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction.
651 *ISME J.* **4**: 673–685.

652 Mulligan, C., Fischer, M., and Thomas, G.H. (2011) Tripartite ATP-independent periplasmic (TRAP)
653 transporters in bacteria and archaea. *FEMS Microbiol. Rev.* **35**: 68–86.

654 Penn, K., Wang, J., Fernando, S.C., and Thompson, J.R. (2014) Secondary metabolite gene expression and
655 interplay of bacterial functions in a tropical freshwater cyanobacterial bloom. *ISME J.* **8**: 1866–1878.

656 Pinhassi, J. and Berman, T. (2003) Differential growth response of colony-forming alpha- and gamma-
657 proteobacteria in dilution culture and nutrient addition experiments from Lake Kinneret (Israel), the
658 eastern Mediterranean Sea, and the Gulf of Eilat. *Appl. Environ. Microbiol.* **69**: 199–211.

659 Poretsky, R.S., Sun, S., Mou, X., and Moran, M.A. (2010) Transporter genes expressed by coastal
660 bacterioplankton in response to dissolved organic carbon. *Environ. Microbiol.* **12**: 616–627.

661 Pruesse, E., Peplies, J., and Glöckner, F.O. (2012) SINA: Accurate high-throughput multiple sequence
662 alignment of ribosomal RNA genes. *Bioinformatics* **28**: 1823–1829.

663 Quinlan, A.R. (2014) BEDTools: The Swiss-Army tool for genome feature analysis. *Curr. Protoc.*
664 *Bioinforma.* **2014**: 11.12.1–11.12.34.

665 Raes, J. and Bork, P. (2008) Molecular eco-systems biology: towards an understanding of community
666 function. *Nat. Rev. Microbiol.* **6**: 693–699.

667 Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential
668 expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–40.

669 Rousseeuw, P.J. (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J.*
670 *Comput. Appl. Math.* **20**: 53–65.

671 Saito, M.A., McIlvin, M.R., Moran, D.M., Goepfert, T.J., DiTullio, G.R., Post, A.F., and Lamborg, C.H.
672 (2014) Multiple nutrient stresses at intersecting Pacific Ocean biomes detected by protein biomarkers.
673 *Science* **345**: 1173–7.

674 Satinsky, B.M., Crump, B.C., Smith, C.B., Sharma, S., Zielinski, B.L., Doherty, M., et al. (2014)

Microspatial gene expression patterns in the Amazon River Plume. *Proc. Natl. Acad. Sci.* **111**: 11085–90.

Shilova, I.N., Robidart, J.C., James Tripp, H., Turk-Kubo, K., Wawrik, B., Post, A.F., et al. (2014) A microarray for assessing transcription from pelagic marine microbial taxa. *ISME J.* **8**: 1476–1491.

Sjöstedt, J., Koch-Schmidt, P., Pontarp, M., Canbäck, B., Tunlid, A., Lundberg, P., et al. (2012) Recruitment of members from the rare biosphere of marine bacterioplankton communities after an environmental disturbance. *Appl. Environ. Microbiol.* **78**: 1361–1369.

Smith, D.P., Thrash, J.C., Nicora, C.D., Lipton, M.S., Burnum-Johnson, K.E., Carini, P., et al. (2013) Proteomic and transcriptomic analyses of “Candidatus Pelagibacter ubique” describe the first PII-independent response to nitrogen limitation in a free-living alphaproteobacterium. *MBio* **4**: e00133-12.

Teeling, H. et al. (2012) Substrate-Controlled Succession of Marine Bacterioplankton Populations Induced by a Phytoplankton Bloom. *Science* (80-.). **336**: 608–611.

Tringe, S.G., von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., et al. (2005) Comparative metagenomics of microbial communities. *Science* (80-.). **308**: 554–557.

Won, N. Il, Kim, K.H., Kang, J.H., Park, S.R., and Lee, H.J. (2017) Exploring the impacts of anthropogenic disturbance on seawater and sediment microbial communities in Korean coastal waters using metagenomics analysis. *Int. J. Environ. Res. Public Health* **14**: 130. doi: 10.3390/ijerph14020130.

Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Qian, G., et al. (2012) FastUniq: A Fast De Novo Duplicates Removal Tool for Paired Short Reads. *PLoS One* **7**: e0052249.

Supporting information

Fig. S1. Map of the sampling and experiment sites. DOMhum site - Lappväärti river in Finland used to collect pristine humic rich river water, DOMagri site – Leilupe river in Latvia influenced by agricultural pollution. Exp I (Helsingør) experiment site with Øresund microbial community, Exp II (Tvärminne) experiment site with microbial community from Storfjärden, Gulf of Finland.

Fig. S2. Total DOC consumption in Exp I (left) and Exp II (right). The DOC consumption was calculated from O₂ consumption based on oxygen measurements (details in *Experimental procedures S1*).

Fig. S3. Inorganic nutrient concentrations in microcosms from Exp I (left panel) and Exp II (right panel). Ammonium concentration is shown in upper panel, nitrate concentration in the middle panel and phosphate concentration in the bottom panel. Details of significant differences between treatments and controls (Dunnett’s test) can be seen in Table S1. DOMhum, fresh water from humic rich river; DOMagri, fresh water from polluted river influenced by agriculture; DOMsea, autochthonous dissolved organic matter; LTC, cell lysate from cyanobacterial culture; NP, inorganic nutrient addition; LowO₂, reduced O₂ concentration; Control, controls.

Fig. S2. Schematic illustration of the experimental setup performed at two sites in the Baltic Sea. Treatments

were performed in triplicates and controls as six replicates. DOMhum, fresh water from humic rich river; DOMagri, fresh water from polluted river influenced by agriculture; DOMsea, autochthonous dissolved organic matter; LTC, cell lysate from cyanobacterial culture; NP, inorganic nutrient addition; LowO₂, reduced O₂ concentration; Control, controls. A detailed description of the experiment is provided in *Experimental procedures*.

Fig. S4. Taxonomic richness and diversity comparison. A) Abundance-based coverage estimator (ACE) of bacterial communities in treatments from Exp I (red) and Exp II (blue) of triplicate samples. B) Shannon index of bacterial communities in treatments from Exp I (red) and Exp II (blue) of triplicate samples. DOMhum, fresh water from humic rich river; DOMagri, fresh water from polluted river influenced by agriculture; DOMsea, autochthonous dissolved organic matter; LTC, cell lysate from cyanobacterial culture; NP, inorganic nutrient addition; LowO₂, reduced O₂ concentration

Fig. S5. MDS plot of treatments from Exp I (circles) and Exp II (triangles) based on COG functional profiles. Distances correspond to leading log₂-fold-changes (logFC) between each sample. Controls (black), DOMagri (yellow), DOMhum (red), DOMsea (orange), LowO₂ (purple), LTC (green), NP (blue).

Fig. S6. COGs assigned to selected major functional categories as a percent of total number of significant COGs displaying higher abundance compared to controls in a) Exp I and b) Exp II. The ‘Transporter and metabolism’ category sums percent COGs related to Amino acid transport and metabolism (E), Carbohydrate transport and metabolism (G), Coenzyme transport and metabolism (H), Energy production and conversion (C), Inorganic ion transport and metabolism (P), Lipid transport and metabolism (I), Nucleotide transport and metabolism (F), Secondary metabolites biosynthesis, transport and catabolism (Q). The ‘Others’ category sums percent COGs related to Cytoskeleton (Z), Extracellular structures (W), Chromatin structure and dynamics (B), RNA processing and medication (A), Posttranslational modification, protein turnover, chaperones (O) and not annotated (NA). Data represent average from triplicate microcosms.

Fig. S7. Venn diagrams of overlapping significant COGs between treatments from Exp I and Exp II either up represented or down represented relative to controls. The overlapping COGs for DOMhum were COG0235, COG1638, COG3772, COG1346, COG1014, COG4231, COG2828, COG5410, COG2358, COG4666, COG4626, COG3740, COG4695, COG5545, COG1783, COG5511, COG5525, COG5323 and COG0234 (up represented) and COG1494 and COG1376 (down represented). The overlapping COGs for DOMagri were COG2195 (down represented) and COG0222 (up represented). For DOMsea overlapping COGs were COG2849, COG2195, COG1629, COG5520, COG3537, COG2356 (up represented) and COG1690 (down represented). LTC showed a huge overlap and an additional filter was applied resulting in 19 overlapping COGs with 9 up represented (COG2819, COG3508, COG4152, COG2849, COG4704, COG4874, COG3757, COG1368 and COG1979) and 10 down represented (COG1180, COG2401, COG5126, COG1276, COG2519, COG2368, COG0638, COG1405, COG5059 and COG5534). The overlapping COGs of NP treatment were COG3772 and COG1783 (up represented) and no overlapping COG down represented. No overlapping COG was identified for treatment LowO₂. See Table S2 for annotations.

Fig. S8. Schematic illustration of the experimental setup performed at two sites in the Baltic Sea. Treatments were performed in triplicates and controls as six replicates. DOMhum, fresh water from humic rich river; DOMagri, fresh water from polluted river influenced by agriculture; DOMsea, autochthonous dissolved organic matter; LTC, cell lysate from cyanobacterial culture; NP, inorganic nutrient addition; LowO₂, reduced O₂ concentration; Control, controls. A detailed description of the experiment is provided in *Experimental procedures*.

767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782

Table S1. Dunnett’s test.

Table S2. COGs that were statistically ($p < 0.01$; $FDR > 0.05$) up represented in experimental treatments when compared to the control microcosms. COGs are listed after additional quality control step (details in *Experimental procedures*). LogFC are listed relative to controls and k-medoids cluster are indicated as number 1-5 (see Fig. 3).

Table S3. Up represented DOC-related COGs and transporters.

Table S4. Association between measured bulk variables/rates and abundance of significantly up or down represented COGs. COGs were clustered using DCA analysis (see also Fig. 5) and bulk variables/rates fitted linearly with COG scores.