Evaluation of a eukaryote phylogenetic microarray for environmental monitoring of marine sediments

Katrine Lekang[1,2*], Anders Lanzén[3,4], Inge Jonassen[5], Eric Thompson[1,6,7], Christofer Troedsson[1,7,8]

[1]Department of Biology, University of Bergen, Bergen, Norway

[2]Department of Pharmacy, University of Oslo, Norway

[3]AZTI-Tecnalia, Marine Research Division, Pasaia, Spain

[4]IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain

[5]Computational Biology Unit, Department of Informatics, University of Bergen, Norway

[6]Sars International Centre for Marine Molecular Biology, University of Bergen, Bergen, Norway

[7]NORCE, Bergen, Norway

[8]Ocean Bergen AS, Bergen, Norway

*Corresponding author: Katrine Lekang

Address: Department of pharmacy, UIO, Sem Sælands vei 3, 0371 Oslo, Norway

Phone: (+47) 93862536

E-mail: Katrine.Lekang@gmail.com

Running title: Microarray for monitoring marine sediments

Abstract

Increased exploitation of resources in sensitive marine ecosystems emphasizes the importance of knowledge regarding ecological impacts. However, current bio-monitoring practices are limited in terms of target-organisms and temporal resolution. Hence, developing new technologies is vital for enhanced ecosystem understanding. In this study, we have applied a prototype version of a phylogenetic microarray to assess the eukaryote community structures of marine sediments from an area with ongoing oil and gas drilling activity. The results were compared with data from both sequencing (metabarcoding) and morphology-based monitoring to evaluate whether microarrays were capable of detecting ecosystem disturbances. A significant correlation between microarray data and chemical pollution indicators, as well as sequencing-based results, was demonstrated, and several potential indicator organisms for pollution-associated parameters were identified, among them a large fraction of microorganisms not covered by traditional morphology-based monitoring. This suggests that microarrays have a potential in future environmental monitoring.

FUNDING: This work was supported by the Research Council of Norway [190265/S40] and Equinor.

DECLARATION OF INTERESTS: The authors have declared that no competing interests exist

65 # Introduction

66 The ocean provides valuable resources such as food, energy and materials. Harvesting these
67 resources can substantially impact marine ecosystems. Current knowledge regarding the
68 bioecological effects of anthropogenic activities, such as oil extraction, trawling and deep-sea
69 mining is limited. Marine ecosystems contain a complex network of interacting organisms
70 (Arrigo, 2005), yet it is only larger, visible organisms that are commonly considered in
71 monitoring programs, despite evidence that microorganisms play key roles in maintaining
72 ecosystem functions (Bik et al., 2012). Therefore, to better understand anthropogenic impacts
73 on an ecosystem, a more complete diversity of organisms needs to be considered.

74 To investigate the effects of petroleum exploitation, microscopy-based monitoring programs
75 with taxonomic classification of macro- and to some extent meiofauna are conducted on
76 benthic samples (Diaz et al., 2004; Gray, 2000; Miljødirektoratet, 2015). This is time-
77 consuming and does not allow frequent assessments of samples (Baird and Hajibabaei, 2012;
78 Brodin et al., 2012; Hajibabaei et al., 2011). Because of this, sampling schedules are often
79 conducted with long temporal intervals, e.g. every third year (OSPAR, 2007), limiting the
80 capacity to distinguish between anthropogenic short and long term effects (e.g. of oil drilling
81 and climate change) and natural factors. It is therefore of interest to develop new, more
82 efficient methods to generate ecosystem data in environmental samples, such as marine
83 sediments (Baird and Hajibabaei, 2012; Chariton et al., 2010; Leray and Knowlton, 2015).

84 Previous studies have suggested that the implementation of molecular high throughput
85 methods could improve biological monitoring (Aggelen et al., 2010; Baird and Hajibabaei,
86 2012; Brodin et al., 2012; Gescher et al., 2008; Hajibabaei et al., 2011; Lallias et al., 2015;
87 Lanzén et al., 2016; Leray and Knowlton, 2015; Thomsen and Willerslev, 2015). The use of
88 DNA based methods makes it possible to include microorganisms in the assessments (Lallias

et al., 2015) and allows us to obtain information on taxa affiliated with several trophic levels in a biological system (Lanzén et al., 2016). This enables more complete assessment of anthropogenic impacts on the ecosystem and provides insight into impacts on ecosystem structure, beyond binary ("affected"/"not affected"). Molecular high throughput assays, using sequencing or microarray hybridization of phylogenetic marker genes, can provide more objective analyses when samples from many locations are compared and when conducting environmental monitoring over long periods of time (Baird and Hajibabaei, 2012), since these methods are less subjective to errors from morphometric assessments by individual taxonomists (Mann et al., 2010). They can also increase the rate and cost-effectiveness of sample processing (Ansorge, 2009; Wetterstrand, 2012). Metagenomic sequencing does indeed provide more information and is able to obtain a deeper characterization of genomes and microbial communities compared to microarrays. However, when optimized, microarrays may serve as an attractive tool for routine, more targeted monitoring of a high number of samples, with both costs and time benefits (Thissen et al., 2019). Microarrays also have the potential for implementation as part of automatic remote sensing pipelines such as an Environmental Sample Processor (ESP), where samples can be collected and processed in situ, with direct data transfer to land for analysis (Jones et al., 2008; Preston et al., 2009). This would be advantageous in routine monitoring of remote areas with limited infrastructure, e.g. deep-sea habitats and areas covered with ice. Automatization may also be beneficial from an economical perspective, since it can reduce boat time and therefore significant costs during these monitoring programs. These benefits can allow for increased temporal resolution, potentially allowing detection of early warning signals, preceding state changes in ecosystems associated with negative and often permanent alterations in ecosystem functioning (Scheffer et al., 2012).

Several studies have tested the potential of high throughput molecular methods, sequencing in particular, for monitoring environments and ecosystem health (Caldwell Eldridge et al., 2017; Carew et al., 2013; Hajibabaei et al., 2011; Kisand et al., 2012; Lanzén et al., 2016; Lejzerowicz et al., 2015). Further, microarrays targeting algae that cause toxic blooms have been tested and integrated in environmental monitoring (Diercks et al., 2008; Dittami et al., 2013a; Dittami et al., 2013b; Edvardsen et al., 2013; Galluzzi et al., 2011). Also, microarrays targeting a broader diversity have been designed for the 16S small subunit (SSU) ribosomal RNA (rRNA) gene and tested on environmental samples (DeSantis et al., 2007; Dubinsky et al., 2012; Nemir et al., 2010; Wang et al., 2017; Yergeau et al., 2009; Zhao et al., 2017). However there are several challenges related to using microarrays for assessing environmental samples (Avarre et al., 2007; Zhou and Thompson, 2002). High sample complexity has been demonstrated to decrease hybridization specificity (Koltai and Weingarten-Baror, 2008) and quantification problems arise because of PCR biases (Palmer et al., 2006; Taberlet et al., 2012).

In this study, a previously developed microarray design (Lekang et al., 2018) was tested on sediment samples from an ongoing environmental monitoring program on the Norwegian continental shelf, to evaluate the impact on targeted taxa by several environmental and discharge parameters. The results were compared to data obtained by metabarcoding (Lanzén et al., 2016) and morphology-based monitoring (DNV, 2011). The main objective was to evaluate the potential of integrating phylogenetic microarrays in routine monitoring using this prototype version of a phylogenetic microarray.

Methods

Samples

Sediment grab samples were collected by Det Norske Veritas (DNV) and Molab as part of an environmental monitoring program in the North Sea, Region III in May 2010 (DNV, 2011), using a van Veen grabb. Aliquots of 50-100 g of sediment were transferred to 250 ml plastic containers (Kautex Textron) and fixed using 96% ethanol, to a final concentration of 70-80%. Samples were stored at -20 °C until further processing. Sediment properties, such as geographical position, grain size and content of chemical compounds were assessed and reported by DNV and Molab (DNV, 2011). In total, 30 samples were included in this study. The fields included were Oseberg C (OSEC); station 05, 06, 08-10 and 15-18, Oseberg D (OSED); station 01, 03-05 and 08, and Veslefrikk (VFR); station 01-11, 20-21, K1-K3 (Figure S1). These were selected based on chemical and physical properties of the samples, which established gradients optimal for such an assessment. Physical parameters (depth and distance from platform), sediment characteristics (grain size, composition; sand, silt/clay and gravel) and chemical parameters (Total Organic Material (TOM), Total Hydrocarbons (THC), Polycyclic Aromatic Hydrocarbons (PAH), Napthtalene Phenantren and Dibenzothiophene (NPD), Barium (Ba), Cadmium (Cd), Chromium (Cr), Copper (Cu), Mercury (Hg), Lead (Pb), Zinc (Zn)) are listed in Table S1. The same samples were also studied by metabarcoding (Lanzén et al., 2016) and by traditional morpho-taxonomic techniques (DNV, 2011).

Sample preparation for microarray analysis

To assess whether we could obtain biological data relevant to environmental monitoring using microarray analyses, genomic DNA extracted from sediments sampled from the 30 samples

are described in the previous section (Table S1). In this study, we used the same genomic DNA extracted for the sequencing analysis (Lanzén et al., 2016). Briefly, genomic DNA was extracted in 10 replicates of 0.5 g sediment from each sample using the PowerSoil® DNA extraction kits (MO BIO Laboratories Inc., Carlsbad CA) (Lekang et al., 2015). The replicate genomic DNA extracts were pooled prior to quantification and PCR. Genomic DNA extracts were quantified using a Qubit® 2.0 Fluorometer (Invitrogen). PCR amplification targeting 18S SSU rRNA was carried out using 25 µl Hot Start Taq Master Mix (Qiagen) and 1 µg/µl of Bovine Serum Albumin (BSA, Thermo Scientific). To each PCR reaction, 2.5 µl of the template was added. To each reaction, 0.5 µM of each of the primers, F-566 and R-1200 (Hadziavdic et al., 2014) with a T7-promotor attached to the reverse-primer (R-1200-T7) were used. PCR amplification was carried out in a thermal cycler (C1000TM Thermal Cycler, BioRad) using the following program: 95 °C for 15 min, 35 cycles consisting of 95 °C for 45 sec, 60 °C for 45 sec, 72 °C for 1 min, and a final extension step of 72 °C for 10 min. Ten replicate PCR reactions were run per sample. Amplification was verified with gel electrophoresis. Positive PCR products were pooled and purified using Agencourt AMPure XP (Beckman Coulter Inc).

From each sample, 500 ng of the PCR product was used as template in the RNA transcription reaction using the MEGAscript T7 kit (Ambion) following the manufacturer's protocol, with the exception that 5-(3-Aminoallyl)-UTPs (InvitrogenTM) was included in a 1:1 ratio to UTP. Five replicate transcriptions for each sample were conducted. The reactions were incubated at 37 °C for 4 hours. RNA transcripts were pooled and purified in two replicates with MEGAclearTM Transcription Clean-Up Kit (Ambion), -precipitation with 5 M Aluminum Acetate and eluted in 25 µl nuclease free water. Replicates for each of the samples were pooled and the final samples quantified by Qubit.

185 The RNA was labeled with Cy3 Mono-Reactive Dye Pack (Amersham), following
186 manufacturer's recommendations. To each labeling reaction, 10 µg RNA was added to a tube
187 of Cy3 dye. The labeling reaction was stopped using 8 µl of 1M Tris-EDTA, pH 8 (Sigma).
188 Labeled RNA was further purified using MEGAclearTM Transcription Clean-Up Kit
189 (Ambion) to eliminate excess Cy3-molecules. Both staining and purification were conducted
190 in an ozone-free environment, and the Cy3-labeled RNA was quantified using NanoDrop®
191 ND-1000 Spectrophotometer. Labeled RNA was split in aliquots of 4 replicates, stored at -80
192 °C and further fragmented and hybridized within 5 days.

193

194 ***Microarray experiment***

195 The labeled RNA was hybridized using a previously designed and optimized microarray
196 (V.1.2) (Lekang et al., 2018). In this specific microarray, the probes were designed to target
197 208 OTUs obtained from a metabarcoding assessment (Lanzén et al., 2016), of the sediment
198 samples included in this study. This strategy, where the probes target OTUs rather than
199 taxonomic group, makes it possible to also detect undescribed organisms. This is an
200 advantage since such organisms represent a substantial fraction of the benthic biodiversity.
201 Each OTU was targeted by several unique probes to reduce the risk of false positives. The
202 process of designing probes has been described in detail, in a previous study (Lekang et al.,
203 2018).

204 The samples were hybridized in replicates of four. The replicates were randomly distributed
205 among the microarrays. For each sample replicate, 50 ng RNA was used for hybridization.
206 Labeled RNA was eluted to a final volume of 19 µl using nuclease free water, according to
207 the protocol from the manufacturer (Agilent technologies). Then, 5 µl of 10X blocking agent
208 (Agilent) and 1 µl of 25X fragmentation buffer (Agilent) were added to each reaction and

incubated at 60 °C for 30 minutes. The fragmentation reaction was stopped by placing the samples on ice for 1 min. Before hybridizations, 25 µl of 2X GE hybridization buffer HI-RPM (Agilent) was added to each reaction and centrifuged 1 min at 13 000 rpm. Finally, 40 µl of the hybridization mixture was loaded onto gasket slide wells and the microarray slides were placed on top with probes facing down. The arrays were hybridized at 61 °C in a rotating oven for 17 hours. After hybridization, slides were washed using a Gene expression wash buffer kit (Agilent) following manufacturer's recommendations. Scanning was performed immediately after washing using an Agilent G2505B (Agilent Technologies). Fragmentation, hybridization, wash and scanning of slides were conducted in an ozone free environment.

Data analysis

Data was extracted from microarray images using Feature Extraction v. 10.7.3.1 (Agilent Technologies) and imported to the Software J-Express 2012 build 119 (Dysvik and Jonassen, 2001). Several filters were applied in J-Express to remove spots flagged by the feature extraction software due to pixel variation (glsFeatNonUnifOL and glsBGNonUnifOL), outlier status compared to replicate probes (glsFeatPopnOL and glsBGPopnOL), background noise (lsWellAboveBG) or saturated spots (glsSaturated) as calculated in the feature extraction step. Median values were calculated for replicate probes on each array and the 4 replicates of each sample were quantile normalized. The data was stored in CSV format. To filter the data we used a six-step filtration pipeline (Lekang et al., 2018) in order to decrease false positives caused by cross-hybridization. Briefly described, the filtration removed OTUs that did not obtain a satisfactory signal in a certain number of probes, and further normalized over-estimated intensity values of probes due to cross-hybridization. We filtered all data with both

233 average filtration (all replicas together) and individual filtration (individual replicas) as
234 previously described (Lekang et al., 2018).

235 Statistical analysis was conducted using the R software (R_Development_Core_Team, 2008).
236 Technical variation between hybridization replicates was calculated and compared with
237 sample variation from each of the tree fields. A heatmap was generated using log-transformed
238 data from the 30 sediment samples from VFR, OSEC and OSED using the R-packages vegan
239 (Oksanen et al., 2013) and gplots (Warnes et al., 2015). The information regarding OTU
240 taxonomy was obtained from the previously published metabarcoding study (Lanzén et al.,
241 2016) (Table S2).

242 Microarray data from the sediment samples at VFR, OSEC and OSED were compared to
243 previously published data from microscopy and metabarcoding (DNV, 2011; Lanzén et al.,
244 2016). Initially, Spearman correlation coefficients were calculated to compare relative
245 abundances of sequences obtained by metabarcoding and corresponding hybridization
246 intensity signals. In this analysis, all OTUs targeted in the microarray and further detected by
247 metabarcoding were included. Hellinger transformation was then conducted on the microarray
248 data and on relative abundance data from microscopy and metabarcoding. Bray-Curtis
249 dissimilarity matrices were calculated based on the transformed values, and the matrices were
250 used to perform multivariate statistics tests, conducted using the vegan-package in R
251 (Oksanen et al., 2013). Specifically, non-metric multidimensional scaling (NMDS; function
252 *metaMDS*), permutational ANOVA (PERMANOVA; function a*donis)*, Mantel and partial
253 Mantel-tests were performed. Correlations of environmental parameters to the NMDS
254 coordinates were investigated using the function *envfit*. PERMANOVA was carried out by
255 only including parameters significantly correlated to NMDS coordinates ($p < 0.05$).
256 Parameters were added sequentially, starting with the one with highest correlation to the
257 NMDS coordinates and subsequently removed from the model unless found to be significant

by PERMANOVA. To assess the effect of environmental parameters and diversity profiles, Mantel and Partial Mantel tests were performed for all sediments collectively and for each field (VFR, OSEC and OSED) separately. In both the PERMANOVA and Mantel tests, a separation was made between parameters such as sediment characteristic and depth and parameters associated with contamination.

To identify possible indicator-OTUs from the microarray dataset, Spearman rank correlations between hybridization intensity signals and environmental parameters were determined. A p-value cut-off of 0.05 after Bonferroni correction was applied. Variation between replicate hybridizations from the sample was compared to variation between replicate hybridizations of different samples, by non-parametric comparison of distribution of Bray-Curtis dissimilarities (Wilcoxon Rank Sum Test).

280 # Results

281 ***Evaluation of microarray hybridization and metabarcoding***

282 There was low variation between replicate hybridizations of the 30 samples from Oseberg C,
283 Oseberg D, and Veslefrikk, and the variation between separate samples was demonstrated as
284 higher than replicate variation for individual samples ($p < 3E\text{-}16$; Figure 1).

285 A Mantel-test was performed to compare biodiversity-profiles obtained by microarrays and
286 metabarcoding, and demonstrated significant correlation between the two methods ($r = 0.47$, p
287 $= 0.001$). The quantitative estimates obtained by microarray and metabarcoding were
288 compared, including only OTUs targeted by the microarray, by calculating Spearman rank
289 correlation coefficients across all 30 samples, as well as for the three individual fields (Table
290 1). Relative abundance estimates obtained by the two methods were also correlated, resulting
291 in coefficients ranging from 0.40 –0.63.

292 In total 208 OTUs were included on the microarray tested in this study. When comparing the
293 hybridization data to sequencing data, using only these 208 OTUs, the microarray detected 37
294 – 100% (80% on average) of the OTUs detected by metabarcoding (Table S3). The OTUs
295 detected by metabarcoding (Lanzén et al., 2016), but not by microarray had a relatively low
296 relative mean abundance ($5x10^{-4}$) according to the sequencing results. Among these, 44%
297 were singletons in their respective sample. The OTUs detected by both metabarcoding, and
298 the microarray, had a higher mean relative abundance ($4x10^{-3}$) according to the sequencing
299 results. The number of OTUs not detected by metabarcoding, but detected by the microarray,
300 ranged from 26 in OSEC-06 to 80 in OSEC-08.

301 ***Correlation to environmental parameters***

302 NMDS was performed based on hybridization intensity signals from samples collected at all
303 three fields, along with metabarcoding from the previous study (Lanzén et al., 2016) and

morpho-taxonomy results (DNV, 2011) from the same sample stations (Figure 2). Among the samples from VFR, three samples representing the least contaminated sites within this field (VFR-02, VFR-08 and VFR-11; see Table S1) appear closely together and distinct from other, more contaminated VFR samples (Figure 2a).

Based on NMDS, microarray-based community results were more strongly correlated to most of the environmental parameters (included Barium), compared to what was observed for the metabarcoding-results (Table 2). However, grain size and depth correlated more strongly with morpho-taxonomy based results. NMDS correlations were consistent with PERMANOVA, indicating a significant impact of Barium ($p<0.001$) and depth ($p<0.001$) on community structure. When controlling for depth, a significant impact was still indicated for Barium ($p<0.001$).

According to Mantel-tests, microarray-based community data correlated more strongly to physical, non-contamination related parameters (depth, sand and grain size; $r = 0.33$, $p < 0.001$) than to contaminants ($r = 0.15$, $p < 0.05$; see Table 3). Mantel-tests performed individually on VFR results generated results that were consistent with this ($r = 0.42$ and, $p = 0.002$ for physical parameters *vs*. $r = 0.30$ and $p < 0.05$ for contaminant-related). A partial mantel test discounting influences of depth, sand, and grain size did not confirm a significant influence on community composition of contamination alone when performed on samples from all three fields. Nevertheless, partial Mantel-tests on samples from VFR did indicate a significant influence of contamination ($r = 0.5$, $p < 0.05$), as opposed to results of OSEC or OSED.

Biodiversity and indicator analysis

Hybridization intensity signals for all OTUs detected in the sediment samples are presented in a heatmap with OTUs affiliated to taxonomic groups and a dendrogram presenting the results

of a hierarchical clustering analysis based on Bray-Curtis dissimilarities (Figure 3). Six of the samples from VFR (03, 20, 05, 04, K3 and K1), representing the most contaminated (in terms of Ba and THC) formed a cluster in the dendrogram presented above the heatmap. Finally, all samples from OSED, grouped together according to the hierarchical clustering.

Several OTUs were present in most samples and did not seem to decrease in abundance with a high level of contamination (Figure 3). However, some OTUs and taxa were more abundant in certain samples; e.g. OTUs from the class Cnidaria had high hybridization intensity signals in several samples from OSEC. This was also the case for a cluster of OTUs assigned to Annelida, more specifically, the family Canalipalpata. These OTUs were detected in some VFR samples, but were not correlated to contamination. Several OTUs from Arthropoda and a cluster of OTUs assigned to Ciliophora (Alveolata) were less abundant in samples from VFR, which had high levels of Ba and THC. There were also some OTUs that appeared more abundant in samples from VFR with a high level of Barium and hydrocarbons: OTU16294, assigned to Peridiniales (Dinophyceae, Alveolata) and two OTUs assigned to Ascomycota (Fungi). The two latter OTUs were detected in OSEC-08 and 09, two of the most Ba-rich samples in this field (Table S1).

Several potential indicator OTUs were identified based on correlations of hybridization intensity signals with environmental parameters (Table S4). These OTUs were taxonomically assigned to Metazoa (5 OTUs), Alveolata (4 OTUs) and Fungi (2 OTUs). Of the metazoan OTUs, four were assigned to Arthropoda and one to Gastrotricha. Two of the OTUs from Alveolata were assigned to Dinophyceae and the other two OTUs to Ciliophora. Both of the fungal OTUs were affiliated with Ascomycota. Three OTUs correlated with parameters describing depth or sediment characteristics (OTU20507; Gastrotricha, OTU21201; Ciliophora, OTU8414; Ciliophora). Most of the correlations to environmental parameters were negative. However, both fungal taxa were positively correlated with Ba, whereas both

353 Dinophyceae taxa were positively correlated with THC, Ba and Hg. Additionally, three OTUs
354 that could not be taxonomically classified (OTU20507, OTU21201 and OTU8414), had a
355 positive correlation with sand.
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372

Discussion

In this study, an 18S rRNA microarray (Lekang et al., 2018) was used to evaluate phylogenetic microarrays as a method for environmental monitoring of marine sediments. This method was applied to a set of samples also analyzed using microscopy (DNV, 2011) and metabarcoding (Lanzén et al., 2016), and the resulting data were compared. Furthermore, biological aspects, such as distribution of taxonomic groups in the benthos and possible indicator OTUs for oil and gas drilling contamination, were assessed.

Evaluation of the microarray assessment

Comparisons between microarray datasets and results obtained by sequencing have demonstrated correlations between the two methods (Brodie et al., 2006; Tottey et al., 2013; Yergeau et al., 2009), in agreement with our results. Several OTUs, not present in the metabarcoding dataset, were detected only by the microarray (Table S3). These might either represent true diversity in the samples, not detected by sequencing, or false positives. Several studies comparing microarrays and metabarcoding have indicated significant correlations when using higher taxonomical levels, such as phyla and class (Claesson et al., 2009; van den Bogert et al., 2011; Yergeau et al., 2009). However, correlation typically decreases with more resolved taxonomic levels (e.g. family or genus) (Claesson et al., 2009; van den Bogert et al., 2011), indicating cross-hybridization between closely related taxa. Here we used a more taxonomy-independent approach, instead based on probes chosen from individual OTUs defined by *de novo* clustering of metabarcoding data. Nonetheless, using the same microarray design but from a previous study, we demonstrated that several false positive hybridizations correspond to OTUs with high sequence similarity to true positive OTUs, and that some OTUs were classified within the same genus as true positive OTUs (Lekang et al., 2018). This

suggests that many false positives may be explained by cross-hybridization to closely related species or strains. Thus, the microarray biodiversity profiles provide meaningful biological information because changes in biodiversity patterns will be reflected in the microarray results, although the presence of specific strains may be challenging without further optimization of the microarray.

Reproducibility is critically important for biodiversity studies since data are compared spatially and temporally, and large variation within samples will generate noise that limits this comparison. Microscopy-based investigations of environmental samples depend on manual evaluations of morphological features by individual taxonomists. The data might therefore vary depending on the person conducting the survey (Archibald, 1984; Mann et al., 2010; Morales et al., 2001) especially at higher taxonomic resolution. Variation in quality might therefore impact conclusions on long time-series. Molecular methods based on phylogenetic marker genes are considered to be more objective, because taxonomical identification is done by comparing nucleic acid sequences (Zimmerman et al., 2014). Still, it is important to assess variation between replicate samples as well as the reproducibility of molecular methods. In this study, four technical hybridization replicates were included for each sample. The VFR, OSEC and OSED samples exhibited low variation between replicates (Figure 1), and variation between hybridization replicates was significantly lower than variation between different samples. These results demonstrate that the microarray is able to distinguish biodiversity signals between separate samples.

Correlation to environmental parameters

Microarray technology has previously been proposed as a tool with good potential for environmental monitoring (Rich, 2011; Rivas, 2011; Wang et al., 2017). Here we aimed to assess this by comparing microarray-based results to morpho-taxonomy and metabarcoding

results from the same samples. A fully developed microarray or metabarcoding approach may potentially provide information on all taxonomic groups, including microorganisms. This is an advantage because smaller organisms quickly respond to changes in the environment due to their small size and rapid generation time (Santos et al., 2010). Furthermore, metabarcoding is more universal in the sense that it can cover all organisms targeted by the primers used, whereas microarrays are restricted to specific taxa targeted by the probes. Compared to morpho-taxonomy techniques, however, molecular methods are not directly quantitative, but rather semi-quantitative, because quantitative abundances can primarily be assessed between samples or over time of the same taxa but not strictly between taxa in one sample (D'Amore et al., 2016). However, changes in biological composition relative to environmental parameters are arguably more important than the number of individuals from each taxonomic group.

In this study, both metabarcoding and morphology data yielded better separation of sites as compared to the microarray data. The microarray-based diversity profiles obtained in this study correlated equally well to most environmental parameters tested, particularly to those associated with contamination or disturbance, such as Barium (Table 2). The metabarcoding results did not correlate as strongly with these parameters. However, higher correlation has been demonstrated by splitting the sequence dataset into metazoan and non-metazoan sequences in a previous study based on the same metabarcoding dataset (Lanzén et al., 2016) (data not shown here). Sample VFR05 was indicated to be most affected by contaminants according to both morphology and microarray results (Figure 2). Indeed, VFR05 was also the most contaminated sample in reference to chemical data (Table S1). This suggested that results from microarray and microscopy yielded similar conclusions based on correlations between community composition and contaminants.

Even though a positive correlation was demonstrated between all contaminants and the community structure profiles obtained by the microarray, this was likely an effect of

447 autocorrelation between contaminant levels rather than suggesting a biological effect from all
448 contaminants. Out of the measured disturbance indicators, Barium was indicated as the most
449 strongly correlated to community structure, according to PERMANOVA, which agrees well
450 with practices and experiences of current monitoring. Barium is a heavy metal, often used as
451 an indicator of drilling activity because it is a component of barite (BaSO4), which is present
452 in drilling mud (Breuer et al., 2004), drill fluids, and other fluids used in offshore petroleum
453 activities (Neff et al., 1987).

454 Based on the microarray results, depth, sand and grain size appeared to affect community
455 structure stronger than the chemical parameters. However, this seems to vary among the oil
456 fields because the effects of contaminants were suggested to be more significant for VFR
457 compared to OSEC and OSED (Table 3).

458

459 ***Indicator organisms***

460 Environmental parameters correlated significantly with the abundance of several taxa, thus
461 potentially useful of indicator organisms in routine monitoring (Table S4). This included four
462 OTUs assigned to Copepoda, negatively correlated to several contaminants, including Ba and
463 Pb. Copepoda has previously been demonstrated to be highly sensitive to petroleum
464 contamination (Bonsdorff, 1981; Frithsen et al., 1985). Two OTUs assigned to Macrodasyida
465 and Euplotida (Hypotrichia) were also negatively correlated to depth and Cd. These two taxa
466 were also sensitive to these environmental parameters in the metabarcoding study (Lanzén et
467 al., 2016). Interestingly, a positive correlation to sand content (%) was also demonstrated with
468 both microarray and sequencing for both of these taxa. Several organisms are known to
469 respond positively to pollution (Frithsen et al., 1985). In this study, two OTUs assigned to
470 Microascales (Fungi), correlated positively with Ba, which was also supported in the

metabarcoding study (Lanzén et al., 2016). Microascales has been suggested to degrade aromatic hydrocarbons, such as toluene (Prenafeta-Boldu´ et al., 2006). Organisms within Dinophyceae have previously been suggested to ingest hydrocarbons (Cooper, 1968) and this has been demonstrated for Dinophyceae; *Noctiluca scintillans* and *Gyrodinium spirale* (Almeda et al., 2014). After a large oil spill in the Bay of Biscay in 1967 (Torrey Canyon), dinoflagellate blooms were linked with the elimination of crude oil (Cooper, 1968). In this study, two OTUs assigned to Dinophyceae (Alveolata) were demonstrated to correlate positively with THC, Ba and Hg.

An important advantage of methods based on phylogenetic markers, such as 18S rRNA genes, is that we can obtain information from a broader spectrum of organisms in an ecosystem than classical microscopy methods, which mainly focus on macro- and meiofauna. Several studies have previously suggested that stressor-effects on microorganisms differ from larger organisms (Danovaro et al., 1995; Lanzén et al., 2016; Santos et al., 2010). In our study, most organisms included on the microarray were multicellular organisms (metazoans, >80%). Nonetheless, more than 50% of the potential indicator OTUs identified were microorganisms, indicating that these are valuable in environmental monitoring and that changes in their abundance should be considered and included in monitoring programs.

Conclusions

This study demonstrates that our previously developed phylogenetic microarray design is capable of profiling eukaryotic community structure with an accuracy similar to metabarcoding and morpho-taxonomy approaches. Specifically, obtained results were significantly correlated with environmental parameters, including contaminants from offshore oil and gas activities. In combination with high throughput sequencing, microarrays have the

potential to increase the temporal and spatial resolution of environmental monitoring by contributing to a more complete ecosystem understanding of anthropogenic activity effects.

Acknowledgements

This work was supported by the Research Council of Norway (190265/S40) and Statoil. We thank DNV for providing the sediment samples during the routine monitoring cruise in 2010, SAM Marine and all members of the Molecular Ecology research group at NORCE as well as Marc Frischer and Christoffer Schander for valuable discussions during the development of this work. We thank Kjell Petersen at CBU, Christine Stansberg and Rita Holdhus at NMC for technical support.

References

Aggelen, G.V., Ankley, G.T., Baldwin, W.S., Bearden, D.W., Benson, W.H., Chipmann, J.K., Collette, T.W., Craft, J.A., Denslow, N.D., Embry, M.R., Falciani, F., George, S.G., Helbing, C.C., Hoekstra, P.F., Iguchi, T., Kagami, Y., Katsiadaki, I., Kille, P., Liu, L., Lord, P.G., McIntyre, T., O'Neill, A., Osachoff, H., Perkins, E.J., Santos, E.M., Skirrow, R.C., Snape, J.R., Tyler, C.R., Versteeg, D., Viant, M.R., Volz, D.C., Williams, T.D., Yu, L., 2010. Integrating Omic Technologies into Aquatic Ecological Risk Assessment and Environmental Monitoring: Hurdles, Achievments, and Future Outlook. Environmental health perspectives 118, 1-5.

Almeda, R., Connelly, T.L., Buskey, E.J., 2014. Novel insight into the role of heterotrophic dinoflagellates in the fate of crude oil in the sea, Sci Rep, p. 9 pp.

Ansorge, W.J., 2009. Next-generation DNA sequencing techniques. New Biotechnology 25, 195-203.

Archibald, R.E.M., 1984. Diatom illustrations—an appeal. Bacillaria 7, 173-178.

Arrigo, K.R., 2005. Marine Microorganisms and global nutrient cycles. Nature 437, 349-355.

Avarre, J.-C., de Lajudie, P., Béna, G., 2007. Hybridization of genomic DNA to microarrays: A challenge for the analysis of environmental samples. Journal of microbiological methods 69, 242-248.

Baird, D.J., Hajibabaei, M., 2012. Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. Mol Ecol 21, 2039-2044.

Bik, H., Porazinska, D.L., Creer, S., Caporaso, J.G., Knight, J.R., Thomas, W.K., 2012. Sequencing our way towards understanding global eukaryotic biodiversity. Trends Ecol Evol 27, 233-243.

Bonsdorff, E., 1981. The Antonio Gramsci oil spill Impact on the littoral and benthic ecosystems Marine pollution bulletin 12, 301-305.

Breuer, E., Stevenson, A.G., Howe, J.A., Carroll, J., Shimmield, G.B., 2004. Drill cutting accumulations in the Northern and Central North Sea: a review of environmental interactions and chemical fate. Mar Pollut Bull 32, 727-733.

Brodie, E.L., DeSantis , T.Z., Joyner, D.C., Baek, S.M., Larsen, J.T., Andersen, G.L., Hazen, T.C., Richardson, P.M., Herman, D.J., Tokunaga, T.C., Wan, J.M., Firestone, M.K., 2006. Application of a High-Density Oligonucleotide Microarray Approach To Study Bacterial Population Dynamics during Uranium Reduction and Reoxidation. Appl Environ Microbiol 72, 6288-6298.

Brodin, Y., Ejdung, G., Strandberg, J., Lyrholm, T., 2012. Improving environmental and biodiversity monitoring in the Baltic Sea using DNA barcoding of Chironomidae (Diptera). Mol Ecol Res 13, 996-1004.

Caldwell Eldridge, S.L., Driscoll, C., Dreher, T.W., 2017. Using high-throughput DNA sequencing, genetic fingerprinting, and quantitative PCR as tools for monitoring bloom-forming and toxigenic cyanobacteria in Upper Klamath Lake, Oregon, 2013 and 2014, Scientific Investigations Report, Reston, VA.

Carew, M.E., Pettigrove, V.J., Metzling, L., Hoffmann, A.A., 2013. Environmental monitoring using next generation sequencing: rapid identification of macroinvertebrate bioindicator species. Front Zool 10, 1-15.

Chariton, A.A., Batley, G.E., Hardy, C.M., 2010. Improving aquatic ecological assessments. Integr Environ Assess Manag 6, 187-188.

Claesson, M.J., O'Sullivan, O., Wang, Q., NikkilÃ¤, J., Marchesi, J.R., Smidt, H., de Vos, W.M., Ross, R.P., O'Toole, P.W., 2009. Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine, PLoS One, p. e6669.

Cooper, L.H.N., 1968. Scientific consequences of the wreck of the 'Torrey Canyon'. Helgoländer wissenschaftliche Meeresuntersuchungen 17, 340-355.

D'Amore, R., Ijaz, U.Z., Schirmer, M., Kenny, J.G., Gregory, R., Darby, A.C., Shakya, M., Podar, M., Quince, C., Hall, N., 2016. A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling, BMC Genomics.

Danovaro, R., Fabiano, M., Vincx, M., 1995. Meifauna response to the Agip Abruzzo oil spill in subtidal sediments of the Ligurian Sea. Mar Pollut Bull 30, 133-145.

566 DeSantis , T.Z., Brodie, E.L., Moberg, J.P., Zubieta, I.X., Piceno, Y.M., Andersen, G.L., 2007. High-
567 Density Universal 16S rRNA Microarray Analysis Reveals Broader Diversity than Typical Clone Library
568 When Sampling the Environment. Microbial Ecology 53, 371-383.
569 Diaz, R.J., Solan, M., Valente, R.M., 2004. A review of approaches for classifying benthic habitats and
570 evaluating habitat quality. J Environ Manage 73, 165 - 181.
571 Diercks, S., Metfies, K., Medlin, L.K., 2008. Development and adaptation of a multiprobe biosensor
572 for the use in a semi-automated device for the detection of toxic algae. Biosens Bioelectron 23, 1527-
573 1533.
574 Dittami, S.M., Hostyeva, V., Egge, E.S., Kegel, J.U., Eikrem, W., Edvardsen, B., 2013a. Seasonal
575 dynamics of harmful algae in outer Oslofjorden monitored by microarray, qPCR, and microscopy. Env
576 sci Pollut res 20, 6719–6732.
577 Dittami, S.M., Pazos, Y., Laspra, M., Medlin, L., 2013b. Microarray testing for the presence of toxic
578 algae monitoring programme in Galicia (NW Spain). Environ sci Pollut Res 20, 6778-6793.
579 DNV, 2011. Miljøovervåkning Region III 2010 Det Norske Veritas, p. 286p.
580 Dubinsky, E.A., Esmaili, L., Hulls, J.R., Cao, Y., Griffith, J.F., Andersen, G.L., 2012. Application of
581 Phylogenetic Microarray Analysis to Discriminate Sources of Fecal Pollution. Environ Sci Technol 46,
582 4340-4347.
583 Dysvik, B., Jonassen, I., 2001. J-Express: exploring gene expression data using Java Bioinformatics 17,
584 369-370.
585 Edvardsen, B., Dittami, S.M., Groben, R., Brubak, S., Escalera, L., F., R., Reguera, B., Chen, J., Medlin,
586 L., 2013. Molecular probes and microarrays for the detection of toxic algae in the genera Dinophysis
587 and Phalacroma (Dinophyta). Environ sci Pollut Res 20, 6733-6750.
588 Frithsen, J.B., Elmgren, R., Rudnick, D.T., 1985. Responses of benthic meiofauna to long-term, low-
589 level additions of No. 2 fuel oil. Marine Ecology progress series 23, 1-14.
590 Galluzzi, L., Cegna, A., Casabianca, S., Penna, A., Saunders, N., Magnani, M., 2011. Development of an
591 oligonucleotide microarray for the detection and monitoring of marine dinoflagellates. J Microbiol
592 Meth 84, 234-242.
593 Gescher, C., Metfies, K., Frickenhaus, S., Knefelkamp, B., Wiltshire, K.H., Medlin, L.K., 2008. Feasibility
594 of Assessing the Community Composition of Prasinophytes at the Helgoland Roads Sampling Site
595 with a DNA Microarray. Appl. Environ. Microbiol. 74, 5305-5316.
596 Gray, J.S., 2000. The measurement of marine species diversity, with an application to the benthic
597 fauna of the Norwegian continental shelf J Exp Mar Bio Ecol 250, 23-49.
598 Hadziavdic, K., Lekang, K., Lanzén, A., Jonassen, I., Thompson, E.M., Troedsson, C., 2014.
599 Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. PLoS One
600 9, e87624.
601 Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.A.C., Baird, D.J., 2011. Environmental Barcoding: A
602 Next-Generation Sequencing Approach for Biomonitoring Applications Using River Benthos, PLoS
603 One.
604 Jones, W.J., Preston, C.N., Marin III, R., Scholin, C.A., Vrijenhoek, R.C., 2008. A robotic molecular
605 method for in situ detection of marine invertebrate larvae. Mol Ecol Res 8, 540-550.
606 Kisand, V., Valente, A., Lahm, A., Tanet, G., Lettieri, T., 2012. Phylogenetic and Functional
607 Metagenomic Profiling for Assessing Microbial Biodiversity in Environmental Monitoring. PLoS One 7,
608 e43630.
609 Koltai, H., Weingarten-Baror, C., 2008. Specificity of DNA microarray hybridization: characterization,
610 effectors and approaches for data correction. Nucleic Acids Res 36, 2395-2405.
611 Lallias, D., Hiddink, J.G., Fonseca, V.G., Gaspar, J.M., Sung, W., Neill, S.P., Barnes, N., Ferrero, T., Hall,
612 N., Lambshead, J.D., Packer, M., Thomas, W.K., Creer, S., 2015. Environmental metabarcoding reveals
613 heterogeneous drivers of microbial eukaryote diversity in contrasting estuarine ecosystems. ISME J 9,
614 1208-1221.
615 Lanzén, A., Lekang, K., Jonassen, I., Thompson, E.M., Troedsson, C., 2016. High-throughput
616 metabarcoding of eukaryotic diversity for environmental monitoring of offshore oil drilling activities.
617 Mol Ecol DOI:10.1111/mec.13761.

Lejzerowicz, F., Esling, P., Pillet, L., Wilding, T., Black, K.D., Pawlowski, J., 2015. High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems, Sci Rep, p. 10p.
Lekang, K., Hadziavdic, K., Sandned Skaar, K., Jonassen, I., Thompson, E.M., Troedsson, C., 2018. Development and testing of an 18S rRNA phylogenetic microarray for marine sediments. Journal of microbiological methods 154, 95-106.
Lekang, K., Thompson, E.M., Troedsson, C., 2015. A comparison of DNA extraction methods for biodiversity studies of eukaryotes in marine sediments. Aquatic Microbial Ecology 75, 15-25.
Leray, M., Knowlton, N., 2015. DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. Proc Natl Acad Sci USA 112, 2076-2081.
Mann, D., Sato, S., Trobajo, R., Vanormelingen, P., Souffreau, C., 2010. DNA barcoding for species identification and discovery in diatoms. Cryptogam Algol 31, 557–577.
Miljødirektoratet, 2015. Environmental Monitoring of petroleum activities on the Norwegian continental shelf. Norwegian Environment Agency.
Morales, E.A., Siver, P.A., Trainor, F.R., 2001. Identification of diatoms (Bacillariophyceae) during ecological assessments: comparison between Light Microscopy and Scanning Electron Microscopy techniques. Proceedings of the Academy of Natural Sciences of Philadelphia 151, 95–103.
Neff, J.M., Rabalais, N.N., Boesch, D.F., 1987. Offshore oil and gas development activities potentially causing long-term environmental effects, in: Boesch, D.F., Rabalais, N.N. (Eds.), Long-Term Environmental Effects of Offshore Oil and Gas Development. Elsevier Applied Science, London, pp. 149-174.
Nemir, A., David, M.M., Perrussel, R., Sapkota, A., Simonet, P., Monier, J.-M., Vogel, T.M., 2010. Comparative phylogenetic microarray analysis of microbial communities in TCE-contaminated soils. Chemosphere 80, 600-607.
Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2013. vegan: Community Ecology Package.
OSPAR, 2007. Assessment of the impact on the marine environment of offshore oil and gas activity - an overview of monitoring results in the United Kingdom, the Netherlands and Norway. OSPAR commission, London, UK, pp. 1-7.
Palmer, C., Bik, E.M., Eisen, M.B., Eckburg, P.B., Sana, T.R., Wolber, P.K., Relman, D.A., Brown, P.O., 2006. Rapid quantitative profiling of complex microbial populations. Nucleic Acids Res 34, e5.
Prenafeta-Boldu´, F.X., Summerbell, R., de Hoog, G.S., 2006. Fungigrowing onaromatic hydrocarbons: biotechnology's unexpected encounterwith biohazard? femS Microbiol Rev 30, 109-130.
Preston, C.M., Marin III, R., Jensen, S.D., Feldman, J., Birch, J.M., Massion, E.I., DeLong, E.F., Suzuki, M., Wheeler, K., Scholin, C.A., 2009. Near real-time, autonomous detection of marine bacterioplankton on a coastal mooring in Monterey Bay, California, using rRNA-targeted DNA probes. Environ Microbiol 11, 1168-1180.
R_Development_Core_Team, 2008. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.
Rich, V.I., Pham, V.D., Eppley, J., Shi, Y., DeLong, F., 2011. Time-series analyses of Monterey Bay coastal microbial picoplankton using a 'genome proxy' microarray. Environ Microbiol 13, 116-134.
Rivas, L.A., Aguirre, J., Blanco, Y., González-Toril, E., Parro, V., 2011. Graph-based deconvolution analysis of multiplex sandwich microarray immunoassays: applications for environmental monitoring. Environ Microbiol 13, 1421-1432.
Santos, H.F., Cury, J.C., Carmo, F.L., Rosando, A.S., Peixoto, R.S., 2010. 18S rDNA Sequences from Microeukaryotes Reveal Oil Indicators in Mangrove Sediment. PLoS One 5, e12437.
Scheffer, M., Carpenter, S.R., Lenton, T.M., Bascompte, J., Brock, W., Dakos, V., van de Koppel, J., van de Leemput, I.A., Levin, S.A., van Nes, E.H., Pascual, M., Vandermeer, J., 2012. Anticipating Critical Transitions. Science 338, 344-348.
Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., Willerslev, E., 2012. Towards next-geneartion biodiversity assessment using DNA metabarcoding. Molecular Ecology 21, 2045-2050.

669 Thissen, J.B., Be, N.A., McLoughlin, K., Gardner, S., Rack, P.G., Shapero, M.H., Rowland, R.R.R., Slezak,
670 T., Jaing, C.J., 2019. Axiom Microbiome Array, the next generation microarray for high-throughput
671 pathogen and microbiome analysis. PLOS ONE 14, e0212045.
672 Thomsen, P.F., Willerslev, E., 2015. Environmental DNA – An emerging tool in conservation for
673 monitoring past and present biodiversity. Biological Conservation 183, 4-18.
674 Tottey, W., Denonfoux, J., Jaziri, F., Parisot, N., Missaoui, M., Hill, D., Borrel, G., Peyretaillade, E.,
675 Alric, M., Harris, H.M.B., Jeffrey, I.B., Claesson, M.J., O'Toole, P.W., Peyret, P., Brugere, J.-F., 2013.
676 The Human Gut Chip ''HuGChip'', an Explorative Phylogenetic Microarray for Determining Gut
677 Microbiome Diversity at Family Level, PLoS One, p. e62544.
678 van den Bogert, B., De Vos, W.M., Zoetendal, E.G., Kleerebezem, M., 2011. Microarray Analysis and
679 Barcoded of Pyrosequencing Provide Consistent Microbial Profiles Depending on the Source Human
680 Intestinal Samples. Appl Environ Microbiol 77, 2071-2080.
681 Wang, J., Zhao, W., Liang, Y., Han, Y., Wang, X., Zhang, X., Ma, X., Zhang, D., Huang, Z., 2017.
682 Development of a prototype 16S rRNA gene-based microarray for monitoring planktonic
683 actinobacteria in shrimp ponds. International Aquatic Research 9, 71-80.
684 Warnes, G.S., Bolker, b., Bonebakker, l., Gentleman, R., Liaw, W.H.A., Lumley, T., Maechler, M.,
685 Magnusson, A., Moeller, S., Schwartz, M., Venables, B., 2015. Package 'gplots': Various R
686 Programming Tools for Plotting Data.
687 Wetterstrand, K., 2012. DNA Sequencing costs. Data from the NHGRI Genome Sequencing Program
688 (GSP)
689 Yergeau, E., Schoondermark-Stolk, S.A., Brodie, E.L., Dejean, S., DeSantis, T.Z., Goncalves, O., Piceno,
690 Y.M., Andersen, G.L., Kowalchuk, G.A., 2009. Environmental microarray analyses of Antarctic soil
691 microbial communities. ISME J 3, 340-351.
692 Zhao, W., Wang, J., Liang, Y., Huang, Z., 2017. Development of a 16S rRNA gene-based microarray for
693 the detection of marine bacterioplankton community. Acta Oceanologica Sinica 36, 106-114.
694 Zhou, J., Thompson, D.K., 2002. Challenges in applying microarrays to environmental studies Curr
695 Opin Biotechnol 13, 204-207.
696 Zimmerman, J., Glõckner, G., Jahn, R., Enke, N., Gemeinholzer, B., 2014. Metabarcoding vs.
697 morphological identification to assess diatom diversity in environmental studies Mol Ecol Res 15,
698 526-542.

699

700

701

702

703

704

705

706

Figure and Table legends

Figure 1: Boxplot presenting the variation among hybridization replicates and between distinct samples from the tree locations; **a)** Oseberg C, **b)** Oseberg D and **c)** Veslefrikk

Figure 2: NMDS based on Bray-Curtis dissimilarities of community composition from **a)** microarray, **b)** metabarcoding using data presented in (Lanzén et al., 2016) and **c)** microscopy using data presented in (DNV, 2011). Hybridization intensities and relative abundances of sequence-reads were Hellinger-transformed and significant environmental parameters marked with blue vectors.

Figure 3: Heatmap of log-transformed hybridization intensity signals from Veslefrikk, OsebergC and OsebergD sediment samples. Color gradients are from black to yellow, representing low to high intensity, respectively. The color on the left axis indicates taxonomic group as coded in the accompanying legend.

Figure S1: Map indicating the geographic position of sampled oilfields. All petroleum fields were included in the metabarcoding analysis (Lanzén et al., 2016), whereas the results created the database for creating probes for the microarray. Petroleum fields marked with a circled dot have been used to test the microarray in this study.

Table 1: Spearman correlation (r) and significance (*** = < 0.001) between relative abundance of sequences and hybridization intensity signals for OTUs targeted by probes and additionally detected by sequencing.

Table 2: Correlation of environmental parameters to NMDS clustering patterns. R^2 values for linear correlation of parameter vectors with maximal correlation to NMDS space resulting from Bray-Curtis distance of Hellinger transformed hybridization intensity signals are displayed together with significance (* = < 0.05, ** = < 0.01, *** = < 0.001) as determined by *envfit* in the R package vegan.

Table 3: Mantel test statistics: permutation-based Mantel tests were used to evaluate the correlation between two dissimilarity matrices ("explanatory" and "dependent" variables below). Bray-Curtis dissimilarity was used to derive community dissimilarities, and log-transformation for environmental parameters.

Table 4: Indicator organisms identified based on significant Spearman correlation with environmental parameters. Significance after Bonferroni correction is indicated by (* = < 0.05, ** = < 0.01, *** = < 0.001).

Table S1: All sediment samples included in this study with the corresponding environmental parameters (DNV, 2011).

Table S2: Taxonomic affiliation of all 264 OTUs targeted by the microarray. The OTUs are indicated by colors based on results from the plasmid experiment (microarray V.1.1); Green = OTU present in the plasmid-mix, green*=false negative, red = false positive, blue = true negative.

Table S3: Sediment samples from Veslefrikk (VFR), OsebergC (OSEC) and OsebergD (OSED) with information on the number of OTUs detected by sequencing, which were targeted by the microarray (1). The table also gives the % positive OTUs based on the microarray (2) and false positive OTUs at each station with reference to the pyrosequencing data (3).

Table S4: Indicator organisms identified based on significant Spearman correlation with environmental parameters. Significance after Bonferroni correction is indicated by (* = < 0.05, ** = < 0.01, *** = < 0.001).