High throughput shotgun sequencing of eRNA reveals taxonomic and derived functional shifts across a benthic productivity gradient

Elias Broman[1], Stefano Bonaglia[1], Alf Norkko[2], Simon Creer[3], and Francisco Nascimento[1]

[1]Stockholm University
[2]University of Helsinki
[3]Bangor University

May 5, 2020

Abstract

Benthic macrofauna is regularly used in monitoring programmes, however the vast majority of benthic eukaryotic biodiversity lies mostly in microscopic organisms, such as meiofauna (invertebrates < 1 mm) and protists, that rapidly responds to environmental change. These communities have traditionally been hard to sample and handle in the laboratory, but DNA sequencing has made such work less time consuming. Compared to DNA sequencing that captures both alive and dead organisms, environmental RNA (eRNA) can be used to better target alive communities. Here, we assessed the biodiversity of three known bioindicator microeukaryote groups (nematodes, foraminifera, and ciliates) in sediment samples collected at seven coastal sites along an organic carbon (OC) gradient. We aimed to investigate if eRNA shotgun sequencing can be used to simultaneously detect differences in 1) biodiversity of multiple microeukaryotic communities, and 2) functional feeding traits of nematodes. Results showed that biodiversity was lower for nematodes and foraminifera in high OC (6.2–6.9 %), when compared to low OC sediments (1.2–2.8 %). The beta diversity for all three groups were different along the OC gradient, as well as the classified feeding type of nematode genera (with more non-selective deposit feeders in high OC sediment). High relative abundant genera included nematode *Sabatieria* and foraminifera *Elphidium* in high OC, and *Cryptocaryon*-like ciliates in low OC sediments. Considering that future sequencing technologies are likely to decrease in cost, the use of eRNA shotgun sequencing to assess biodiversity of living benthic microeukaryotes could be a powerful tool in recurring monitoring programmes.

Introduction

Biodiversity is decreasing globally due to human alteration and pollution of terrestrial and aquatic environments (Brondizio, Settele, Díaz, & Ngo, 2019). Essential ecosystem services affiliated with human health, such as availability of food, clean water, and recreational areas are dependent on biodiversity (Cardinale et al., 2012; Pan, Marcoval, Bazzini, Vallina, & Marco, 2013). In addition to the provision of ecosystem services, biodiversity losses have also been linked to a decrease in ecosystem stability (McCann, 2000). Anthropogenic pressure on coastal aquatic ecosystems by e.g. climate change, eutrophication and contaminant pollution threatens the diversity of many organisms in these systems (Pan et al., 2013). Anthropogenic pressure in coastal ecosystems should be taken seriously because coastal zones are transitional areas directly adjacent to human settlements between land and sea, and impacted areas are predicted to increase in both number and area with a continued climate change scenario (Levin et al., 2001; Rabalais, Turner, Díaz, & Justić, 2009). It is therefore essential to understand how the diversity of organisms living in coastal zones respond to changes in environmental gradients and anthropogenic pressure (Snelgrove, Thrush, Wall, & Norkko, 2014).

Biodiversity assessments of benthic macrofauna are commonly used in national monitoring programs, including coastal zones, to determine various ecological indexes (Pinto et al., 2009). However, microeukaryotes

present in sediment such as meiofaunal nematodes (< 1 mm body size) are also known to react to e.g. eutrophication status (Ristau, Spann, & Traunspurger, 2015), and the composition and quantity of organic carbon (OC) (Ingels, Kiriakoulakis, Wolff, & Vanreusel, 2009; Pusceddu, Gambi, Zeppilli, Bianchelli, & Danovaro, 2009). Furthermore, because nematodes are known to have different feeding behaviours such as bacterivory, detritivory or algal feeding (Moens, Traunspurger, & Bergtold, 2006; Wieser, 1953) changes in nematode assemblages are therefore likely to affect food web dynamics and ecosystem function (e.g. Nascimento et al., 2019; Nascimento, Karlson, & Elmgren, 2008; Nascimento, Näslund, & Elmgren, 2012). Other arguments for including meiofauna such as nematodes in national monitoring systems include their high diversity, short generation time, and ubiquitous distribution (Kennedy & Jacoby, 1999). However, these organisms are often neglected in monitoring studies (Kennedy & Jacoby, 1999), likely due to financial reasons derived from time consuming activities such as sieving, sorting, and microscopic morphological analyses.

In addition, the protist phyla *Foraminifera* (henceforth forams) and *Ciliophora* (i.e. ciliates) are well-studied as bioindicators of environmental state of aquatic ecosystems. The diversity and community composition of forams are known to change with anthropogenic pollution, fish farming, and decreasing water quality (Damak, Frontalini, Elleuch, & Kallel, 2019; Frontalini & Coccioni, 2011; Jan Pawlowski, Esling, Lejzerowicz, Cedhagen, & Wilding, 2014; Raposo et al., 2018; Uthicke & Nobes, 2008), and similar to nematodes, OC enrichment of the sediment also influences the diversity of forams (Martins et al., 2015). Ciliates are used as bioindicators in wastewater treatment plants (Chen, Xu, Tam, Cheung, & Shin, 2008; Foissner, 2016). In natural aquatic environments, the diversity and community composition of ciliates are influenced by e.g. salinity, pH, and anthropogenic pollution (e.g. Gong et al., 2015; Jiang, Xu, Hu, Warren, & Song, 2013). One of the main merits of assessing the diversity of protists as bioindicators is their documented rapid change to environmental conditions (Payne, 2013). The assessment of both meiofaunal and protist biodiversity is therefore a good proxy in monitoring programmes to study changes in ecosystems. However, these communities are rarely studied together and challenges still include being able to investigate multiple communities from bulk sediment samples without time consuming activities involved in studying the benthic meiofaunal fraction such as sieving, sorting, and microscopy.

In the last ten years, environmental DNA (eDNA) metabarcoding studies using the 18S rRNA marker gene have been extensively conducted to study microeukaryotes including nematodes, forams, and ciliates (Bik et al., 2012; Elias Broman et al., 2019; Carugati, Corinaldesi, Dell'Anno, & Danovaro, 2015; Fonseca et al., 2010; Forster et al., 2019; Lallias et al., 2014; Lara & Acosta-Mercado, 2012; Nascimento et al., 2019; J. Pawlowski, Lejzerowicz, & Esling, 2014; Peham, Steiner, Schlick-Steiner, & Arthofer, 2017). Such tools have drastically reduced the time needed to taxonomically classify organisms compared to morphological taxonomic techniques, that also involves sieving and sorting of organisms (Carugati et al., 2015). However, limitations exist with DNA metabarcoding such as non-optimized PCR protocols and primer bias when targeting multiple taxa (Kelly, Shelton, & Gallego, 2019). In addition to using eDNA that will assess the biodiversity of both living organisms plus non-degraded DNA from dead organisms, environmental RNA (eRNA) is targeting only living organisms or RNA derived from organisms of recent origin in the environment (Cristescu, 2019; S. A. Wood et al., 2020). With new bioinformatic tools that can taxonomically classify hundreds of millions of sequences within minutes to hours (e.g. D. E. Wood, Lu, & Langmead, 2019) and estimate relative abundances at species or genus level (e.g. Lu, Breitwieser, Thielen, & Salzberg, 2017), it is valuable to investigate how eRNA combined with the latest sequencing technology can be used to assess differences in biodiversity of active multiple communities from highly diverse and dense environments such as soils and sediments, while avoiding the biases associated with PCR amplification.

Here we assessed the biodiversity and community composition of three microeukaryotic groups in sediment samples: nematodes, forams, and ciliates, along an OC gradient in a coastal archipelago in the Gulf of Finland, Baltic Sea. The aim was to investigate if eRNA shotgun sequencing, without any sieving or sorting of samples (i.e. bulk sediment), could be used to detect differences in biodiversity of multiple microeukaryotic communities. Additionally, we assessed if changes in nematode functional ecology (feeding type) as a response to the OC gradient could be detected. We expected that nematode deposit feeders would have higher relative abundance in stations with higher OC. This approach was coupled to the latest sequencing platform (Illumina

NovaSeq S4, yielding ~87 million sequences per sample), and new bioinformatic tools to estimate taxonomic classifications and relative abundances from data of this size (Kraken 2 + Bracken 2 combination). The Gulf of Finland is characterised by strong environmental gradients associated with eutrophication (Andersen et al., 2015; Villnäs et al., 2019). This contributes to spatially heterogenous benthic macro-communities in terms of diversity and composition in this ecosystem (Bonsdorff, Laine, Hanninen, Vuorinen, & Norkko, 2003). The Gulf of Finland is therefore a well-suited system to investigate if a similar heterogeneity exists in active microeukaryotic communities.

Materials and methods

Field sampling

Sediment was collected on board R/V Electra during 2018 September 20–23 in the Gulf of Finland (Baltic Sea) close to the Tvärminne Zoological Station, Finland. A total of seven stations were visited along coastal gradients in depth and OC (0–4 km, 10–45 m water depth; Figure 1). The stations were divided into four low % OC shallow sites (stations 11, 12, 15, 16; 1.2–2.8 % OC) and three sites with higher % OC and depth (stations 7, 10, 13; 6.2–6.9 % OC), following a station labelling system used during reoccurring monitoring in the Tvärminne region (Table 1). Triplicate sediment cores (labelled A, B, C), retrieved in rinsed acrylic core liners, were collected from each station with a GEMAX twin gravity corer (height: 80 cm, inner diameter: 90 mm). The top 0–2 cm sediment surface layer was sliced into autoclaved 215 ml polypropylene containers (Noax Lab). After slicing, the sediment was directly aseptically homogenized inside the containers and 2 ml sediment transferred into 2 ml cryogenic tubes (VWR) that were immediately flash frozen at -196ºC. The samples were transported on dry ice and stored at -80°C until RNA extraction. The remaining sediment in the 215 ml containers were stored at -20°C for sediment C and N content and pore water chemistry analyses.

Sediment and pore water chemistry analyses

The remaining sediment in the frozen 215 ml containers were thawed, homogenized, and 15 ml sediment from each sample was dried at 60ºC for seven days for C/N analysis. In addition, 20 ml of sediment from each sample was centrifuged at 2200 × g to extract the pore water for ammonium (NH_4^+) and phosphate (PO_4^{3-}) analyses. The dried sediment was ground, homogenized, and 1 ml dry weight sediment per sample stored in a desiccator prior to freeze drying, re-grinding, re-homogenizaton and treated with HCl to remove inorganic carbon. Samples were subsequently weighed into tin capsules. Concentrations of total OC and total nitrogen were determined on an elemental analyser (Flash 2000, Thermo Scientific). The pore water was collected after centrifugation by filtering 10 ml of the supernatant through a 0.45 µm polyethersulfone membrane filter (Filtropur S 0.45, Sarstedt). NH_4^+ and PO_4^{3-} were determined colorimetrically (Multiskan GO spectrophotometer, Thermo Scientific) and NH_4^+ analysis followed the modified salicylate-hypochlorite method by Bower and Holm-Hansen (1980), and PO_4^{3-} analysis followed the standard methods for seawater analyses (Grasshoff, Kremling, & Ehrhardt, 2009).

RNA extraction and sequencing

Sediment was thawed within minutes inside the cryotubes and ~2 g of material was added into the RNeasy PowerSoil bead tubes and was extracted using the same kit (RNeasy PowerSoil, QIAGEN). After RNA extraction, any remaining DNA was removed with DNase treatment using the TURBO DNA-free kit (Invitrogen), followed by bacterial rRNA depletion using the RiboMinus Transcriptome Isolation Kit (bacteria version, ThermoFisher Scientific). Library preparation followed the TruSeq RNA Library Prep v2 kit (Illumina) without including the poly-A selection step. The samples were sequenced at the Science for Life Laboratory, Stockholm on a single Illumina NovaSeq6000 S4 lane using paired-end 2 × 150 bp read technology. A full list of sample names, sequences yielded, quality scores, read lengths etc. are available in Supplementary Data 1.

Bioinformatics

The sequencing yielded on average 87.3 million paired-end sequences per sample (range 77.7–97.8 million sequences). Illumina adapters were removed with SeqPrep 1.2 (St John, 2014) following default settings

with parameters -A and -B targeting the adapter sequences with identical selection. Any remaining PhiX sequences in the raw data were removed by mapping the reads using bowtie2 2.3.4.3 (Langmead & Salzberg, 2012) against the PhiX genome (NCBI Reference Sequence: NC_001422.1). Final quality trimming of the data was conducted with Trimmomatic 0.36 (Bolger, Lohse, & Usadel, 2014) with the following parameters: LEADING:20 TRAILING:20 MINLEN:50. The final quality of the trimmed reads were checked with FastQC 0.11.5 (Andrews, 2010) and MultiQC 1.7 (Ewels, Magnusson, Käller, & Lundin, 2016). On average 86.8 million sequences remained (range 77.3–97.2 million sequences) with a Phred quality score of 36–37 per base, and an average read length of 144 bp (range 139–147 bp).

Small subunit (SSU) 18S rRNA sequences were extracted from the quality trimmed data using SortMeRNA 2.1b (Kopylova, Noé, & Touzet, 2012) with the databases supplied with the software. Taxonomic classification was conducted with Kraken2 2.0.7 (D. E. Wood et al., 2019) using paired-end reads against the SILVA SSU r132 NR99 (Quast et al., 2013) (database downloaded 1 March 2019) and NCBI NT database (database downloaded 12 March 2019). To estimate the relative abundance of each taxa Bracken 2.5 was used on the Kraken2 outputs with default settings on the genus level (i.e. a count threshold of 10) (Lu et al., 2017). The Bracken output reports were combined into a biom-format file with the python package kraken-biom 1.0.1 (using parameters: —fmt hdf5 -max D –min G), and the biom-format file was converted to a tax table using the python package biom-format 2.1.7 (McDonald et al., 2012). The final data was normalized as relative abundances (%), and the data was analyzed in the software Explicet 2.10.5 (Robertson et al., 2013). Results for i)*Nematoda* (NCBI NT classifications, on average 478,331 sequences per sample); ii) *Foraminifera* (NCBI NT, average 13,913 sequences), and iii) *Ciliophora* (SILVA, average 774,027 sequences) were extracted and analyzed separately. The NCBI NT database was used for the *Nematoda* and *Foraminifera* data because, 1) the SILVA database is known to contain errors in the nematode classifications (Elias Broman et al., 2019; Holovachov, Haenel, Bourlat, & Jondelius, 2017), and the NCBI NT has previously been used to discern differences in nematode communities on a spatial scale in the Baltic Sea (Elias Broman et al., 2019); and 2) the SILVA database gave inaccurate classifications for *Foraminifera* , resulting in the identification of taxa never discovered in the Baltic Sea (more details in the discussion).

Nematoda functional ecology analyses

Nematode genera were classified into feeding types based on their known buccal cavity morphology in available literature according to Wieser (1953). Each genus was categorized into the four feeding types described by Weiser: 1A) selective deposit feeder, 1B) non-selective deposit feeder, 2) epistrate feeder, and 2B) predator/omnivore. In addition, the maturity index (MI) of each nematode community was calculated to infer changes in the life history characteristics of nematode genera. MI was calculated according to Bongers, Alkemade, and Yeates (1991) by assigning colonizer–persister (cp) values to nematode genera on a scale from 1 to 5 based in available literature. Low cp-values indicate nematode genera that can be classified as colonizers (short life cycle, high reproduction rates, high colonization ability and tolerance to disturbance) while high cp-values represent persisters (nematode genera that display long life cycles, few offspring, low colonization ability and high sensitivity to disturbance). MI could then be calculated from:

$$MI \; = \; \sum_{i=1}^{n} v\,(i) \times f(i)$$

where $\nu(\imath)$ is the cp-value of genus i and $f(i)$ is the frequency of genus i .

Statistics

Rarefaction curves of sequence counts versus the taxonomic classifications were conducted in the R package vegan (Oksanen et al., 2018) using the *rrarefy* function with default settings. Species richness (Chao1) and alpha diversity (Shannon's H) were calculated in the software Explicet for each taxonomic group (*Nematoda* ,*Foraminifera* , and *Ciliophora*). Before calculating Shannon's H index the data was sub-sampled to the lowest sample size and bootstrapped × 100 (*Nematoda* 79,815 counts, *Foraminifera*2473 counts, *Ciliophora*

299,504 counts). Non-metric multidimensional scaling (NMDS) plots showing beta diversity were based on the presence/absence (Sørensen index) using the software past 3.26 (Hammer, Harper, & Ryan, 2001). The difference in read abundance between the high and low OC stations for *Nematoda* feeding type data was normalized and statistically tested using the R package DESeq2 1.26 with default settings (Love, Huber, & Anders, 2014). The DESeq2 output was plotted using the ggplot2 package in R (Wickham, 2016). Differences between groups on alpha diversity metrics (Chao1, Shannon's H), relative abundance of taxonomic groups, and maturity index for nematodes were tested with univariate statistics conducted in the software IBM SPSS Statistics 26. First, Shapiro-Wilk tests were used to check if the data was normally distributed. Differences between groups in normally distributed data were tested with One-Way ANOVA tests, while non-parametric data were tested with Mann-Whitney U tests. PERMANOVA tests (9999 permutations) were used to identify differences in beta diversity between groups and stations in the software past 3.26. To investigate if the abiotic variables (% OC, % N, PO_4^{3-}, NH_4^+, and water depth) were associated with the community composition canonical correspondence analysis (CCA) was conducted in the R package vegan (Oksanen et al., 2018) with the*cca* function and plotted using the ggplot2 package. The input data for the CCAs were the measured abiotic variables and relative abundances of the different taxa. Significant associations between abiotic variables and community compositions were tested for CCA axis 1 and axis 2 with PERMANOVA tests (9999 permutations) using the function*envfit* included in the vegan package.

Results

Field data and sediment characteristics

Both OC and N content in the sediment were higher at the deeper stations, i.e. $>$ 30 m water depth (6.5 $\pm$ 0.2 % OC and 0.9 $\pm$ 0.02 % N) when compared to the shallow stations, i.e. $<$ 30 m water depth (1.8 $\pm$ 0.1 % OC and 0.2 $\pm$ 0.02 % N; Mann-Whitney U tests, $U = 0$, $P = 0.000007$ for both; Table 1). Hereafter the shallow stations will therefore be referred to as "Low OC" and the deep stations as "High OC". Pore water NH_4^+ and PO_4^{3-} extracted from the top 0–2 cm sediment layer at the seven sampled stations (Table 1 & Figure 1,$n = 3$ for each station) correlated positively with water depth (both $P < 0.01$, Pearson correlations, $r = 0.83$ and $r = 0.64$, respectively). NH_4^+ was significantly higher at High OC stations (308 $\pm$ 19.8 µg/l, n = 12) compared to the Low OC stations (196 $\pm$ 14.3 µg/l (mean $\pm$ SE, $n = 9$; Mann-Whitney U test,U = 10, $P = 0.002$). Similarly, pore water PO_4^{3-} was significantly higher at High OC stations (32.2 $\pm$ 7.0 µg/l) compared to the Low OC (4.4 $\pm$ 0.5 µg/l; Mann-Whitney U test, $U = 1$, $P = 0.000140$). A full list of abiotic data for all stations is available in Supplementary Data 2.

Alpha and beta diversity

Here we studied three microeukaryotic groups: *Nematoda* ,*Foraminifera* , and *Ciliophora* and rarefaction analyses showed that the majority of the genera had been detected in the samples (Supplementary Figure 1). The species richness Chao1 index was lower for*Nematoda* and *Foraminifera* at the High OC stations compared to Low OC (One-Way ANOVA test for each group; *Nematoda* ,$F_{(1,19)} = 32.7$, $P = 0.000016$;*Foraminifera* , $F_{(1,19)} = 57.0$, $P = 0.0000004$; Figure 2a-b). No difference in species richness was observed for *Ciliophora* (Figure 2c). Shannon's H alpha diversity index was also significantly lower for *Nematoda* and *Foraminifera* at the High OC stations when compared to Low OC (One-Way ANOVA test for each group; *Nematoda* , $F_{(1,19)} = 24.8$,$P = 0.000083$; *Foraminifera* , $F_{(1,19)} = 48.2$, $P = 0.000001$; Figure 2d-e). No difference in Shannon's H alpha diversity index was observed for *Ciliophora* (Figure 2f). A full list of Shannon's H values is available in Supplementary Data 3.

Beta diversity was also significantly different between stations for all three groups, with the presence/absence Sørensen index test (PERMANOVA (9999 permutations) test for each group, *Nematoda* , pseudo$F = 4.9$, P = 0.0001; *Foraminifera* , pseudo $F = 10.5$, $P = 0.0001$; *Ciliophora* , pseudo $F = 3.1$,$P = 0.0001$; Figure 3). Similarly, there was a difference in beta diversity when testing the stations grouped as Low OC against High OC (PERMANOVA (9999 permutations) test for each group, *Nematoda* , pseudo $F = 11.4$, $P = 0.0001$; *Foraminifera* , pseudo$F = 25.5$, $P = 0.0001$; *Ciliophora* , pseudo $F = 5.1$, $P = 0.0001$; Figure 3).

CCAs based on the relative abundance of genera showed that the measured abiotic variables (water depth,

sediment % C and % N, plus pore water NH_4^+ and PO_4^{3-}) were associated with the High OC stations for all of the three studied taxonomic groups, i.e.*Nematoda* , *Foraminifera* , and *Ciliophora* (Figure 4). The CCA analysis showed that 67 %, 77 % and 66 % of the total constrained inertia for nematodes, foraminifera and ciliophoran, respectively, was explained with the five environmental variables here studied. There was also a significant association between all five abiotic variables and the community composition for each studied group (PERMANOVA test (9999 permutations), *Nematoda* ,R^2 = 0.76–0.83, $P < 0.001$;*Foraminifera* , R^2 = 0.52–0.94, $P< 0.05$; *Ciliophora* , R^2 = 0.53–0.85, $P < 0.01$; Supplementary Data 4).

Differences in Nematoda community structure

The *Nematoda* 18S rRNA dataset showed differences in community composition along the OC gradient. This included e.g. the nematode genus*Sabatieria* that was detected at all stations and had a significantly higher relative abundance at the High OC stations when compared to Low OC (52.2 ± 7.9 % compared to 22.2 ± 3.5 % (% denote portion of *Nematoda* community, Mann-Whitney U test, $U = 16$, $P = 0.006$; Figure 5a). Similarly, the genus*Axonolaimus* had a higher relative abundance at High OC stations (8.9 ± 2.3 % compared to 0.7 ± 0.2 %, Mann-Whitney U test, $U = 0$, $P = 0.000007$; Figure 5a). In contrast, in the Low OC stations the genera *Daptonema* (19.7 ± 3.0 %) and *Desmolaimus* (4.8 ± 0.7 %) had a higher relative abundance when compared to High OC with*Daptonema* (4.0 ± 0.7 %, $U = 2$, $P = 0.000027$) and*Desmolaimus* (0.7 ± 0.2 %, $U = 1$, $P = 0.000014$) (Mann-Whitney U tests; Figure 5a).

Differences in Foraminifera community structure

Looking closer at *Foraminifera* the genera with a high relative abundance, the genus *Elphidium* had a higher relative abundance among the *Foraminifera* at the High OC stations (66.3 ± 2.7 %) compared to Low OC (16.2 ± 2.1 % (% denote portion of*Foraminifera* community); Mann-Whitney U test, $U = 0$,$P = 0.000007$; Figure 5b). On the other hand, the genus*Rhizammina* had a higher relative abundance at the Low OC stations (17.9 ± 3.1 %) when compared to High OC (1.0 ± 0.8 %; Mann-Whitney U test, $U = 2$, P = 0.000027; Figure 5b).

Differences in Ciliophora community structure

Examples of *Ciliophora* genera that were significant along the OC gradient, included *Cryptocaryon* that had a higher relative abundance at the Low OC stations (17.4 ± 1.4 %) when compared to High OC (12.4 ± 1.4 % (% denote portion of *Ciliophora* community), Mann-Whitney U tests, $U = 21$, $P = 0.018$; Figure 5c). Instead, the genus *Spirotrachelostyla* had a higher relative abundance at High OC (2.9 ± 0.5 %) when compared to Low OC (0.8 ± 0.1 %, Mann-Whitney U test, $U = 9$, $P = 0.00066$; Figure 5c). A full list of taxonomic classifications and sequence counts for all three studied groups are available in Supplementary Data 5.

Nematoda functional ecology

The maturity index calculated from classified *Nematoda* genera showed no difference between High OC and Low OC stations (1.9 ± 0.1 maturity index for all samples, One-Way ANOVA test, Supplementary Data 6). Considering that values closer to one indicate habitat colonizers (and values closer to five indicate habitat persisters) the nematode communities in this study are considered colonizers. Looking closer at the classified feeding type of the nematodes the Genera classified as non-selective deposit feeders (1B, following the classification systems by Wieser (1953)) had a higher number of reads in the High OC stations when compared to Low OC (log2 fold change 1.79, DESeq2 analysis, false discovery rate (FDR) = 0.0000000000018; Figure 6). In contrast, the Low OC stations had more genera classified as selective deposit feeders (1A, log2 fold change 1.62) and predator/omnivores (2B, log2 fold change 1.40) ($FDR = 0.000012$ and $FDR = 0.010$, respectively; Figure 6). A full list of all maturity index and feeding type classifications and their relative abundance per *Nematoda*genera is available in Supplementary Data 6.

Discussion

In this study we investigated if current sequencing technology and eRNA shotgun sequencing has the power

to differentiate changes in biodiversity of multiple microeukaryotes in bulk sediment samples. We focused on nematodes, forams, and ciliates which are useful bioindicators and known to change in diversity and community composition in relation to environmental change (Gong et al., 2015; Ingels et al., 2009; Martins et al., 2015; Jan Pawlowski et al., 2014; Ristau et al., 2015). The results showed a difference in community structure for each of the communities along the OC gradient in the study area. For example, the non-selective deposit feeding nematode genera *Sabatieria* and*Axonolaimus* (Schratzberger, Warr, & Rogers, 2007) had a higher relative abundance at the High OC stations. Potentially this could be a beneficial feeding strategy at the deeper stations where the sediment consists mainly of decayed organic particles and bacteria as food (and is reflected in the nematode feeding type analysis; Figure. 6).*Sabatieria* are typical nematodes found in organic rich sediments, and have been identified with other non-selective deposit feeders such as *Daptonema* (Armenteros et al., 2009; Montagna & Harper, 1996; Schratzberger, Warr, & Rogers, 2006). Interestingly, the genera*Daptonema* and *Desmolaimus* (also a non-selective deposit feeder (Schratzberger et al., 2007)) had a higher relative abundance at the Low OC stations, and they could potentially be directly competing with other nematodes with a similar feeding behavior (such as*Sabatieria*). The Low OC stations had more nematodes classified as selective deposit feeders and predator/omnivores, suggesting different kinds of food (e.g. algae) and more competition in these environments. These nematode genera have previously been detected in other basins the Baltic Sea using 18S rRNA gene metabarcoding (Elias Broman et al., 2019), and here their presence was confirmed by shotgun sequencing.

The foram genera *Elphidium* and *Rhizammina* showed contrasting patterns in the dataset, with *Elphidium* having higher relative abundance at High OC stations, and *Rhizammina* at the Low OC stations. The species *Elphidium excavatum* has previously been found to be dominant in a lagoon located on the Atlantic coast of Portugal with high content of OC (2–4 % OC) compared to adjacent bays with lower OC (Martins et al., 2015). Both*Elphidium* and *Rhizammina* are known to exist in the south-western Baltic Sea (Frenzel, Tech, & Bartholdy, 2005; Schweizer, Polovodova, Nikulina, & Schönfeld, 2011), and to our knowledge, this is the first study using molecular data to investigate diversity of forams in the Gulf of Finland. Even though SILVA is one of the recommended options to classify 18S rRNA sequences of protists (Creer et al., 2016) we were still surprised to see many differences in classified*Foraminifera* genera between the SILVA and NCBI NT databases. For example, SILVA reported a high relative abundance of genera (e.g.*Calcarina* , up to 35% in the offshore stations) never detected in the Baltic Sea (Supplementary Data 5). Almost 100 foram species have been reported from the south-western Baltic Sea, but there is still a large scarcity of studies investigating forams in the central and northern Baltic Sea (Frenzel et al., 2005). It is therefore possible that the differences we observed between databases are due to a limited number of 18S rRNA Baltic Sea-like foram sequences in SILVA. There are specific foram databases such as the PFR^2 (Morard et al., 2015), but this database is focusing on oceanic forams featuring no data from the Baltic Sea (e.g. no data on *Elphidium* and*Rhizammina*). Nevertheless, we report good results derived from the NCBI NT database.

Finally, we did not detect a difference in alpha diversity for ciliates between High OC and Low OC stations. Neither was there a difference between genera with a high relative abundance (except*Cryptocaryon*), and so the difference observed in beta diversity would therefore have been due to differences in low abundant genera. The ciliate genus *Cryptocaryon* was more prominent at the Low OC stations;a marine genus known to include parasitic species targeting fish (Wright & Colorni, 2002). However, low-saline (5–7 ppt) variants of *Cryptocaryon* have previously been described (Yambot, Song, & Sung, 2003). *Cryptocaryon* -like ciliates have previously been detected phylogenetically in the more saline (˜14 ppt) deeper waters of the Baltic Sea (Stock, Jürgens, Bunge, & Stoeck, 2009), and both SILVA and NCBI NT (and manually checking classified sequences via BLAST) confirmed *Cryptocaryon* -like ciliates in our samples. Potentially because the Low OC stations were located in shallow areas closer to the shore, the *Cryptocaryon* -like ciliates detected in this study might be adapted to lower salinities (˜7 ppt) and related to host organisms residing in these more diverse and euphotic habitats. In addition, considering that the summer heatwave of 2018 was one of the most intense ever recorded in the study area (Humborg et al., 2019), the warmer waters might attract*Cryptocaryon* -like ciliates which are typically more common at temperate and tropical temperatures (Colorni & Burgess, 1997). As far as we are aware, this is the first time such species have been reported from the Gulf of Finland based

on molecular data.

Limitations of this study include the relatively small sample size for RNA extraction (2 g, i.e. ˜2 ml for high porosity sediment). A previous study investigating the effect of sample size on diversity estimations of microeukaryotes and metazoans using metabarcoding, found that larger volumes of sediments are necessary to a accurately estimate small-scale spatial heterogeneity in biodiversity (Nascimento, Lallias, Bik, & Creer, 2018). Here we used the available commercial extraction kit that could process the largest input of sediment volume. It will be useful for future studies to develop larger kits for eRNA extraction and investigate similar effects of sample size on biodiversity assessment based on eRNA. Nevertheless, even with this sample size we report clear differences in biodiversity of multiplied communities including metazoans along the environmental gradients. Organisms contain multiple ribosomes and while this is not an issue when analyzing species richness or beta diversity with a presence/absence index, for community composition organisms with a large number of ribosomes could skew the proportions. This issue also exists in DNA metabarcoding studies, with many eukaryotic organisms carrying multiple genome copies per cell (i.e. polyploidy) (Edgar, Zielke, & Gutierrez, 2014) and for prokaryotes that can also be polyploid (Soppa, 2017). The cost of shotgun sequencing on the latest platform is still quite high ($8815 USD per Illumina NovaSeq S4 lane) compared to metabarcoding of marker genes ($1373 USD per Illumina MiSeq flowcell, prices based on the cost at SciLifeLab, Stockholm, Sweden in 2019). However, there has been a large decrease in sequencing cost over the past 20 years (Wetterstrand, 2020) and if this trend continues, alongside streamlined bioinformatic protocols, large scale eRNA shotgun studies could be a future possibility in biomonitoring programmes.

Shotgun sequencing "catches" all organisms in the sample including both prokaryotes and eukaryotes (Zepeda Mendoza, Sicheritz-Pontén, & Gilbert, 2015). Considering that rRNA has a short lifespan in the environment (Blazewicz, Barnard, Daly, & Firestone, 2013) the approach used in this study very likely mostly targeted only living (or very recently dead) organisms, as well as eRNA derived from organisms recently passing through the sampling area but not located in the collected sediment (up to ˜13h or longer in biofilm, see S. A. Wood et al., 2020). This makes eRNA sequencing a useful method to study benthic communities, considering that a substantial portion of sediment consists of long-lasting dead organic matter (Burdige, 2006). For example, forams are a known microfossil group and with the use of DNA extraction it has been possible to study the ancient DNA of these organisms (Lejzerowicz et al., 2013). Shotgun sequencing of eRNA has been used in a wide variety of marine studies, including investigations of prokaryotic communities (Elias Broman, Sachpazidou, Pinhassi, & Dopson, 2017; Elias Broman, Sjöstedt, Pinhassi, & Dopson, 2017; Klindworth et al., 2014; Urich et al., 2014), marine viruses, (Culley, Lang, & Suttle, 2006), sediment eukaryotic metatranscriptomes (Elias Broman, Varvara, Dopson, & Hylander, 2017), and old marine groundwaters in the deep terrestrial biosphere (Lopez-Fernandez et al., 2018). However, there is paucity of studies using eRNA to assess the biodiversity of microeukaryotes in sediment. Although 18S rRNA metabarcoding has gained popularity to investigate such communities (see e.g. Birrer et al., 2018; Comeau, Lagunas, Scarcella, Varela, & Lovejoy, 2019; Rodríguez-Martínez et al., 2020), we have here shown that eRNA shotgun sequencing is also a viable approach to detect differences in diversity and community compositions for multiple communities as a response to different environmental conditions. Even though not directly compared in this study, shotgun sequencing avoids the main limitations of metabarcoding such as 1) PCR primers amplifying species with different rates and 2) the amount of cycles and type of polymerase used influencing diversity and community composition (Kelly et al., 2019; Nichols et al., 2018). In addition, eRNA shotgun studies also provides information on all organisms over a large range of trophic stages in the sediment, and it is also possible to study the RNA transcripts of expressed genes to estimate oxidation and reduction processes of nutrients (mainly related to the metabolism of prokaryotes) (see e.g. Elias Broman et al., 2017). Finally, common shotgun sequencing bioinformatic pipelines are intricate and include many different software which increases the complexity of the data analysis. Here, conversely, we followed a protocol with few and straightforward steps, including: 1) Quality trimming (removal of Illumina adapters and phiX control sequences, quality trimming, and verifying final quality); 2) extraction of SSU rRNA sequences from the dataset; 3) taxonomic classification of the SSU rRNA sequences using Kraken2 against the NCBI NT and SILVA databases; and 4) estimation of relative abundance at genus level using Bracken2 (for more details on the kraken2+bracken2

combination see D. E. Wood et al., 2019). These new bioinformatic tools makes it less daunting and possible to classify large datasets containing hundreds of millions of sequences within minutes to hours which would previously have taken several weeks using traditional aligners. As such this approach is closer to current metabarcoding bioinformatic pipelines with relatively straightforward steps (see e.g. the DADA2 pipeline Callahan et al., 2016).

Conclusions

Here we have shown that eRNA shotgun sequencing is a useful tool to study the biodiversity of benthic microeukaryotes. The latest sequencing technology yields tens of million sequences per sample and this makes it possible to investigate the biodiversity of multiple communities. In our study we focused on three microeukaryotic groups (nematodes, forams, and ciliates). We were able to detect a decrease in biodiversity for nematodes and forams in sediments with higher OC, when compared to low OC sediments. Moreover, we detected differences in beta diversity for all three groups between the stations along the OC gradient, as well as in the functional ecology of nematodes (i.e. feeding type). Considering that future sequencing technologies are likely to develop and decrease in cost, shotgun sequencing of eRNA to assess biodiversity of alive benthic microeukaryotes could be a useful method in recurring monitoring programmes. Taken together, eRNA shotgun sequencing and new bioinformatic tools give the opportunity to study a large diversity of microeukaryotes from bulk sediment samples within a reasonable time frame that was previously not possible.

Acknowledgements

The authors acknowledge support from the National Genomics Infrastructure in Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project SNIC 2018-8-246. Research activities were funded by the Stockholm University's strategic funds for Baltic Sea research. SB was supported by the Swedish Research Council Formas (Grant No. 2017-01513).

Competing interests

We have no competing interests.

References

Andersen, J. H., Carstensen, J., Conley, D. J., Dromph, K., Fleming-Lehtinen, V., Gustafsson, B. G., . . . Murray, C. (2015). Long-term temporal and spatial trends in eutrophication status of the Baltic Sea. *Biological Reviews, 92* (1), 135-149. doi:10.1111/brv.12221

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Armenteros, M., Ruiz-Abierno, A., Fernández-Garcés, R., Pérez-García, J. A., Díaz-Asencio, L., Vincx, M., & Decraemer, W. (2009). Biodiversity patterns of free-living marine nematodes in a tropical bay: Cienfuegos, Caribbean Sea. *Estuarine, Coastal and Shelf Science, 85* (2), 179-189. doi:https://doi.org/10.1016/j.ecss.2009.08.002

Bik, H. M., Porazinska, D. L., Creer, S., Caporaso, J. G., Knight, R., & Thomas, W. K. (2012). Sequencing our way towards understanding global eukaryotic biodiversity. *Trends in Ecology & Evolution, 27* (4), 233-243. doi:10.1016/j.tree.2011.11.010

Birrer, S. C., Dafforn, K. A., Simpson, S. L., Kelaher, B. P., Potts, J., Scanes, P., & Johnston, E. L. (2018). Interactive effects of multiple stressors revealed by sequencing total (DNA) and active (RNA) components of experimental sediment microbial communities. *Science of The Total Environment, 637-638* , 1383-1394. doi:https://doi.org/10.1016/j.scitotenv.2018.05.065

Blazewicz, S. J., Barnard, R. L., Daly, R. A., & Firestone, M. K. (2013). Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *The ISME journal, 7* (11), 2061-2068. doi:10.1038/ismej.2013.102

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics, 30* (15), 2114–2120. doi:10.1093/bioinformatics/btu170

Bongers, T., Alkemade, R., & Yeates, G. W. (1991). Interpretation of disturbance-induced maturity decrease in marine nematode assemblages by means of the Maturity Index. *Marine Ecology Progress Series* , 135-142.

Bonsdorff, E., Laine, A. O., Hanninen, J., Vuorinen, I., & Norkko, A. (2003). Zoobenthos of the outer archipelago waters (N. Baltic Sea)-the importance of local conditions for spatial distribution patterns. *Boreal Environment Research, 8* (2), 135-146.

Bower, C. E., & Holm-Hansen, T. (1980). A Salicylate–Hypochlorite Method for Determining Ammonia in Seawater. *Canadian Journal of Fisheries and Aquatic Sciences, 37* (5), 794-798. doi:10.1139/f80-106

Broman, E., Raymond, C., Sommer, C., Gunnarsson, J. S., Creer, S., & Nascimento, F. J. A. (2019). Salinity drives meiofaunal community structure dynamics across the Baltic ecosystem. *Molecular ecology, 28* (16), 3813-3829. doi:10.1111/mec.15179

Broman, E., Sachpazidou, V., Pinhassi, J., & Dopson, M. (2017). Oxygenation of hypoxic coastal Baltic Sea sediments impacts on chemistry, microbial community composition, and metabolism. *Frontiers in Microbiology, 8* (2453). doi:10.3389/fmicb.2017.02453

Broman, E., Sjöstedt, J., Pinhassi, J., & Dopson, M. (2017). Shifts in coastal sediment oxygenation cause pronounced changes in microbial community composition and associated metabolism. *Microbiome, 5* (96). doi:10.1186/s40168-017-0311-5

Broman, E., Varvara, S., Dopson, M., & Hylander, S. (2017). Diatoms dominate the eukaryotic metatranscriptome during spring in coastal 'dead zone' sediments. *Proceedings of the Royal Society of London B: Biological Sciences, 284* (1864). doi:10.1098/rspb.2017.1617

Brondizio, E., Settele, J., Díaz, S., & Ngo, H. (2019). Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. *IPBES Secretariat* .

Burdige, D. J. (2006). *Geochemistry of marine sediments* : PRINCETON University Press.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nature Methods, 13* (7), 581-583.

Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., . . . Naeem, S. (2012). Biodiversity loss and its impact on humanity. *Nature, 486* (7401), 59-67. doi:10.1038/nature11148

Carugati, L., Corinaldesi, C., Dell'Anno, A., & Danovaro, R. (2015). Metagenetic tools for the census of marine meiofaunal biodiversity: An overview. *Marine Genomics, 24* , 11-20. doi:10.1016/j.margen.2015.04.010

Chen, Q.-H., Xu, R.-L., Tam, N. F. Y., Cheung, S. G., & Shin, P. K. S. (2008). Use of ciliates (Protozoa: Ciliophora) as bioindicator to assess sediment quality of two constructed mangrove sewage treatment belts in Southern China. *Marine Pollution Bulletin, 57* (6), 689-694. doi:https://doi.org/10.1016/j.marpolbul.2008.03.015

Colorni, A., & Burgess, P. (1997). *Cryptocaryon irritans* Brown 1951, the cause of 'white spot disease' in marine fish: an update. *Aquarium Sciences and Conservation, 1* (4), 217-238. doi:10.1023/a:1018360323287

Comeau, A. M., Lagunas, M. G., Scarcella, K., Varela, D. E., & Lovejoy, C. (2019). Nitrate Consumers in Arctic Marine Eukaryotic Communities: Comparative Diversities of 18S rRNA, 18S rRNA Genes, and Nitrate Reductase Genes. *Applied and environmental microbiology, 85* (14), e00247-00219. doi:10.1128/aem.00247-19

Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W. K., . . . Bik, H. M. (2016). The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution, 7* (9), 1008-1018. doi:10.1111/2041-210x.12574

Cristescu, M. E. (2019). Can Environmental RNA Revolutionize Biodiversity Science? *Trends in Ecology & Evolution* . doi:https://doi.org/10.1016/j.tree.2019.05.003

Culley, A. I., Lang, A. S., & Suttle, C. A. (2006). Metagenomic Analysis of Coastal RNA Virus Communities. *Science, 312* (5781), 1795. doi:10.1126/science.1127404

Damak, M., Frontalini, F., Elleuch, B., & Kallel, M. (2019). Benthic foraminiferal assemblages as pollution proxies along the coastal fringe of the Monastir Bay (Tunisia). *Journal of African Earth Sciences, 150* , 379-388. doi:https://doi.org/10.1016/j.jafrearsci.2018.11.013

Edgar, B. A., Zielke, N., & Gutierrez, C. (2014). Endocycles: a recurrent evolutionary innovation for post-mitotic cell growth.*Nature Reviews Molecular Cell Biology, 15* (3), 197-210. doi:10.1038/nrm3756

Ewels, P., Magnusson, M., Käller, M., & Lundin, S. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics, 32* (19), 3047-3048. doi:10.1093/bioinformatics/btw354

Foissner, W. (2016). Protists as bioindicators in activated sludge: Identification, ecology and future needs. *European Journal of Protistology, 55* , 75-94. doi:https://doi.org/10.1016/j.ejop.2016.02.004

Fonseca, V. G., Carvalho, G. R., Sung, W., Johnson, H. F., Power, D. M., Neill, S. P., . . . Creer, S. (2010). Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nature communications, 1* , 98. doi:10.1038/ncomms1095

Forster, D., Filker, S., Kochems, R., Breiner, H.-W., Cordier, T., Pawlowski, J., & Stoeck, T. (2019). A Comparison of Different Ciliate Metabarcode Genes as Bioindicators for Environmental Impact Assessments of Salmon Aquaculture. *Journal of Eukaryotic Microbiology, 66* (2), 294-308. doi:10.1111/jeu.12670

Frenzel, P., Tech, T., & Bartholdy, J. (2005). Checklist and annotated bibliography of Recent Foraminiferida from the German Baltic Sea coast.*Studia Geologica Polonica, 124* , 67-86.

Frontalini, F., & Coccioni, R. (2011). Benthic foraminifera as bioindicators of pollution: A review of Italian research over the last three decades. *Revue de Micropaléontologie, 54* (2), 115-127. doi:https://doi.org/10.1016/j.revmic.2011.03.001

Gong, J., Shi, F., Ma, B., Dong, J., Pachiadaki, M., Zhang, X., & Edgcomb, V. P. (2015). Depth shapes α- and β-diversities of microbial eukaryotes in surficial sediments of coastal ecosystems.*Environmental Microbiology, 17* (10), 3722-3737. doi:10.1111/1462-2920.12763

Grasshoff, K., Kremling, K., & Ehrhardt, M. (2009). *Methods of seawater analysis* : John Wiley & Sons.

Hammer, Ø., Harper, D. A. T., & Ryan, P. D. (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica, 4* (1), 9.

Holovachov, O., Haenel, Q., Bourlat, S. J., & Jondelius, U. (2017). Taxonomy assignment approach determines the efficiency of identification of OTUs in marine nematodes. *Royal Society Open Science, 4* (8). doi:10.1098/rsos.170315

Humborg, C., Geibel, M. C., Sun, X., McCrackin, M., Mörth, C.-M., Stranne, C., . . . Norkko, J. (2019). High Emissions of Carbon Dioxide and Methane From the Coastal Baltic Sea at the End of a Summer Heat Wave. *Frontiers in Marine Science, 6* , 1-14.

Ingels, J., Kiriakoulakis, K., Wolff, G. A., & Vanreusel, A. (2009). Nematode diversity and its relation to the quantity and quality of sedimentary organic matter in the deep Nazaré Canyon, Western Iberian Margin. *Deep Sea Research Part I: Oceanographic Research Papers, 56* (9), 1521-1539.

Jiang, Y., Xu, H., Hu, X., Warren, A., & Song, W. (2013). Functional groups of marine ciliated protozoa and their relationships to water quality. *Environmental Science and Pollution Research, 20* (8), 5272-5280. doi:10.1007/s11356-013-1525-0

Kelly, R. P., Shelton, A. O., & Gallego, R. (2019). Understanding PCR Processes to Draw Meaningful Conclusions from Environmental DNA Studies.*Scientific Reports, 9* (1), 12133. doi:10.1038/s41598-019-48546-x

Kennedy, A. D., & Jacoby, C. A. (1999). Biological Indicators of Marine Environmental Health: Meiofauna – A Neglected Benthic Component?*Environmental Monitoring and Assessment, 54* (1), 47-68. doi:10.1023/A:1005854731889

Klindworth, A., Mann, A. J., Huang, S., Wichels, A., Quast, C., Waldmann, J., . . . Glöckner, F. O. (2014). Diversity and activity of marine bacterioplankton during a diatom bloom in the North Sea assessed by total RNA and pyrotag sequencing. *Marine Genomics, 18* , 185-192. doi:https://doi.org/10.1016/j.margen.2014.08.007

Kopylova, E., Noé, L., & Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data.*Bioinformatics, 28* (24), 3211-3217. doi:10.1093/bioinformatics/bts611

Lallias, D., Hiddink, J. G., Fonseca, V. G., Gaspar, J. M., Sung, W., Neill, S. P., . . . Creer, S. (2014). Environmental metabarcoding reveals heterogeneous drivers of microbial eukaryote diversity in contrasting estuarine ecosystems. *The ISME journal, 9* , 1208. doi:10.1038/ismej.2014.213

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods, 9* , 357. doi:10.1038/nmeth.1923

Lara, E., & Acosta-Mercado, D. (2012). A molecular perspective on ciliates as soil bioindicators. *European Journal of Soil Biology, 49* , 107-111. doi:https://doi.org/10.1016/j.ejsobi.2011.11.001

Lejzerowicz, F., Esling, P., Majewski, W., Szczuciński, W., Decelle, J., Obadia, C., . . . Pawlowski, J. (2013). Ancient DNA complements microfossil record in deep-sea subsurface sediments. *Biology letters, 9* (4), 20130283.

Levin, L. A., Boesch, D. F., Covich, A., Dahm, C., Erséus, C., Ewel, K. C., . . . Weslawski, J. M. (2001). The Function of Marine Critical Transition Zones and the Importance of Sediment Biodiversity.*Ecosystems, 4* (5), 430-451. doi:10.1007/s10021-001-0021-4

Lopez-Fernandez, M., Simone, D., Wu, X., Soler, L., Nilsson, E., Holmfeldt, K., . . . Dopson, M. (2018). Metatranscriptomes reveal that all three domains of life are active but are dominated by bacteria in the Fennoscandian crystalline granitic continental deep biosphere.*mBio, 9* (6), e01792-01718. doi:10.1128/mBio.01792-18

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology, 15* (12), 550. doi:10.1186/s13059-014-0550-8

Lu, J., Breitwieser, F. P., Thielen, P., & Salzberg, S. L. (2017). Bracken: estimating species abundance in metagenomics data. *PeerJ Computer Science, 3* , e104.

Martins, M. V. A., Silva, F., Laut, L. L. M., Frontalini, F., Clemente, I. M. M. M., Miranda, P., . . . Dias, J. M. A. (2015). Response of Benthic Foraminifera to Organic Matter Quantity and Quality and Bioavailable Concentrations of Metals in Aveiro Lagoon (Portugal).*PLoS ONE, 10* (2), e0118077. doi:10.1371/journal.pone.0118077

McCann, K. S. (2000). The diversity–stability debate. *Nature, 405* , 228. doi:10.1038/35012234

McDonald, D., Clemente, J. C., Kuczynski, J., Rideout, J. R., Stombaugh, J., Wendel, D., . . . Meyer, F. (2012). The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. *GigaScience, 1* (1), 7.

Moens, T., Traunspurger, W., & Bergtold, M. (2006). Feeding ecology of free-living benthic nematodes. *Freshwater Nematodes. Ecology and Taxonomy. CAB International Publishing* , 105-131.

Montagna, P., & Harper, J., Donald E. (1996). Benthic infaunal long-term response to offshore production platforms in the Gulf of Mexico. *Canadian Journal of Fisheries and Aquatic Sciences, 53* (11), 2567-2588.

Morard, R., Darling, K. F., Mahé, F., Audic, S., Ujiié, Y., Weiner, A. K. M., . . . de Vargas, C. (2015). PFR2: a curated database of planktonic foraminifera 18S ribosomal DNA as a resource for studies of plankton ecology, biogeography and evolution. *Molecular Ecology Resources, 15* (6), 1472-1485. doi:10.1111/1755-0998.12410

Nascimento, F. J. A., Dahl, M., Deyanova, D., Lyimo, L. D., Bik, H. M., Schuelke, T., . . . Gullström, M. (2019). Above-below surface interactions mediate effects of seagrass disturbance on meiobenthic diversity, nematode and polychaete trophic structure. *Communications Biology, 2* (1), 362. doi:10.1038/s42003-019-0610-4

Nascimento, F. J. A., Karlson, A. M., & Elmgren, R. (2008). Settling blooms of filamentous cyanobacteria as food for meiofauna assemblages. *Limnology and Oceanography, 53* (6), 2636-2643.

Nascimento, F. J. A., Lallias, D., Bik, H. M., & Creer, S. (2018). Sample size effects on the assessment of eukaryotic diversity and community structure in aquatic sediments using high-throughput sequencing. *Scientific Reports, 8* .

Nascimento, F. J. A., Näslund, J., & Elmgren, R. (2012). Meiofauna enhances organic matter mineralization in soft sediment ecosystems. *Limnology and Oceanography, 57* (1), 338-346. doi:doi:10.4319/lo.2012.57.1.0338

Nichols, R. V., Vollmers, C., Newsom, L. A., Wang, Y., Heintzman, P. D., Leighton, M., . . . Shapiro, B. (2018). Minimizing polymerase biases in metabarcoding. *Molecular Ecology Resources, 18* (5), 927-939.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., O'hara, R., Simpson, G. L., . . . Wagner, H. J. h. c. r.-p. o. A. e. (2018). vegan: Community Ecology Package. R package version 2.5-2. In.

Pan, J., Marcoval, M. A., Bazzini, S. M., Vallina, M. V., & Marco, S. (2013). Coastal marine biodiversity: Challenges and threats. In A. H, M. Arias, & C. Menéndez (Eds.), *Marine Ecology in a Changing World*(pp. 43-67): Boca Raton, FL: CRC Press.

Pawlowski, J., Esling, P., Lejzerowicz, F., Cedhagen, T., & Wilding, T. A. (2014). Environmental monitoring through protist next-generation sequencing metabarcoding: assessing the impact of fish farming on benthic foraminifera communities. *Molecular Ecology Resources, 14* (6), 1129-1140. doi:10.1111/1755-0998.12261

Pawlowski, J., Lejzerowicz, F., & Esling, P. (2014). Next-Generation Environmental Diversity Surveys of Foraminifera: Preparing the Future. *The Biological Bulletin, 227* (2), 93-106. doi:10.1086/BBLv227n2p93

Payne, R. J. (2013). Seven reasons why protists make useful bioindicators. *Acta Protozoologica, 52* (3).

Peham, T., Steiner, F. M., Schlick-Steiner, B. C., & Arthofer, W. (2017). Are we ready to detect nematode diversity by next generation sequencing? *Ecology and Evolution, 7* (12), 4147-4151. doi:doi:10.1002/ece3.2998

Pinto, R., Patrício, J., Baeta, A., Fath, B. D., Neto, J. M., & Marques, J. C. (2009). Review and evaluation of estuarine biotic indices to assess benthic condition. *Ecological Indicators, 9* (1), 1-25. doi:https://doi.org/10.1016/j.ecolind.2008.01.005

PRJNA541422. Retrieved from https://www.ncbi.nlm.nih.gov/bioproject/PRJNA541422

Pusceddu, A., Gambi, C., Zeppilli, D., Bianchelli, S., & Danovaro, R. (2009). Organic matter composition, metazoan meiofauna and nematode biodiversity in Mediterranean deep-sea sediments. *Deep Sea Research Part II: Topical Studies in Oceanography, 56* (11), 755-762. doi:https://doi.org/10.1016/j.dsr2.2008.10.012

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., . . . Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research, 41* (Database issue), D590-D596. doi:10.1093/nar/gks1219

Rabalais, N. N., Turner, R. E., Díaz, R. J., & Justić, D. (2009). Global change and eutrophication of coastal waters. *ICES Journal of Marine Science, 66* (7), 1528-1537. doi:10.1093/icesjms/fsp047

Raposo, D., Clemente, I., Figueiredo, M., Vilar, A., Lorini, M. L., Frontalini, F., . . . Laut, L. (2018). Benthic foraminiferal and organic matter compounds as proxies of environmental quality in

a tropical coastal lagoon: The Itaipu lagoon (Brazil). *Marine Pollution Bulletin, 129* (1), 114-125. doi:https://doi.org/10.1016/j.marpolbul.2018.02.018

Ristau, K., Spann, N., & Traunspurger, W. (2015). Species and trait compositions of freshwater nematodes as indicative descriptors of lake eutrophication. *Ecological Indicators, 53* , 196-205. doi:https://doi.org/10.1016/j.ecolind.2015.01.010

Robertson, C. E., Harris, J. K., Wagner, B. D., Granger, D., Browne, K., Tatem, B., . . . Frank, D. N. (2013). Explicet: graphical user interface software for metadata-driven management, analysis and visualization of microbiome data. *Bioinformatics, 29* (23), 3100-3101. doi:10.1093/bioinformatics/btt526

Rodríguez-Martínez, R., Leonard, G., Milner, D. S., Sudek, S., Conway, M., Moore, K., . . . Richards, T. A. (2020). Controlled sampling of ribosomally active protistan diversity in sediment-surface layers identifies putative players in the marine carbon sink. *The ISME journal* . doi:10.1038/s41396-019-0581-y

Schratzberger, M., Warr, K., & Rogers, S. I. (2006). Patterns of nematode populations in the southwestern North Sea and their link to other components of the benthic fauna. *Journal of Sea Research, 55* (2), 113-127. doi:https://doi.org/10.1016/j.seares.2005.07.002

Schratzberger, M., Warr, K., & Rogers, S. I. (2007). Functional diversity of nematode communities in the southwestern North Sea.*Marine Environmental Research, 63* (4), 368-389. doi:https://doi.org/10.1016/j.marenvres.2006.10.006

Schweizer, M., Polovodova, I., Nikulina, A., & Schönfeld, J. (2011). Molecular identification of Ammonia and Elphidium species (Foraminifera, Rotaliida) from the Kiel Fjord (SW Baltic Sea) with rDNA sequences.*Helgoland Marine Research, 65* (1), 1-10. doi:10.1007/s10152-010-0194-3

Snelgrove, P. V. R., Thrush, S. F., Wall, D. H., & Norkko, A. (2014). Real world biodiversity–ecosystem functioning: a seafloor perspective.*Trends in Ecology & Evolution, 29* (7), 398-405. doi:https://doi.org/10.1016/j.tree.2014.05.002

Soppa, J. (2017). Polyploidy and community structure. *Nature Microbiology, 2* , 16261. doi:10.1038/nmicrobiol.2016.261

St John, J. (2014). SeqPrep. In.

Stock, A., Jürgens, K., Bunge, J., & Stoeck, T. (2009). Protistan diversity in suboxic and anoxic waters of the Gotland Deep (Baltic Sea) as revealed by 18S rRNA clone libraries. *Aquatic Microbial Ecology, 55* (3), 267-284.

Urich, T., Lanzén, A., Stokke, R., Pedersen, R. B., Bayer, C., Thorseth, I. H., . . . Øvreas, L. (2014). Microbial community structure and functioning in marine sediments associated with diffuse hydrothermal venting assessed by integrated meta-omics. *Environmental Microbiology, 16* (9), 2699-2710. doi:10.1111/1462-2920.12283

Uthicke, S., & Nobes, K. (2008). Benthic Foraminifera as ecological indicators for water quality on the Great Barrier Reef. *Estuarine, Coastal and Shelf Science, 78* (4), 763-773. doi:https://doi.org/10.1016/j.ecss.2008.02.014

Wetterstrand, K. A. (2020). DNA sequencing costs: data from the NHGRI Genome Sequencing Program (GSP) (Accessed: January 14, 2020). Retrieved from www.genome.gov/sequencingcostsdata

Wickham, H. (2016). *ggplot2: elegant graphics for data analysis* : Springer.

Wieser, W. (1953). Die Beziehung zwischen Mundhohlengestalt, Ernahrungsweise und Vorkommen bei freilebenden marinen Nematoden.*Arkiv for zoologi, 4* , 439–484.

Villnäs, A., Janas, U., Josefson, A. B., Kendzierska, H., Nygård, H., Norkko, J., & Norkko, A. (2019). Changes in macrofaunal biological traits across estuarine gradients: implications for the coastal nutrient filter. *Marine*

Ecology Progress Series, 622 , 31-48.

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology, 20* (1), 257. doi:10.1186/s13059-019-1891-0

Wood, S. A., Biessy, L., Latchford, J. L., Zaiko, A., von Ammon, U., Audrezet, F., . . . Pochon, X. (2020). Release and degradation of environmental DNA and RNA in a marine system. *Science of The Total Environment, 704* , 135314. doi:https://doi.org/10.1016/j.scitotenv.2019.135314

Wright, A.-D. G., & Colorni, A. (2002). Taxonomic re-assignment of*Cryptocaryon irritans* , a marine fish parasite. *European Journal of Protistology, 37* (4), 375-378.

Yambot, A. V., Song, Y.-L., & Sung, H.-H. (2003). Characterization of*Cryptocaryon irritans* , a parasite isolated from marine fishes in Taiwan. *Diseases of aquatic organisms, 54* (2), 147-156.

Zepeda Mendoza, M. L., Sicheritz-Pontén, T., & Gilbert, M. T. P. (2015). Environmental genes and genomes: understanding the differences and challenges in the approaches and software for their analyses.*Briefings in Bioinformatics, 16* (5), 745-758. doi:10.1093/bib/bbv001

Data Accessibility

The raw sequence data have been uploaded and are available on the NCBI database with the following BioProject number PRJNA541422, with an online link listed in the reference section.

Author contributions

EB designed the study, sampled in the field, conducted molecular laboratory work, bioinformatics, analyzed molecular data, and drafted the manuscript. SB conducted chemistry laboratory work, analyzed chemistry data, and gave feedback on the manuscript. AN sampled in the field and gave feedback on the manuscript. SC provided input on the study and feedback on the manuscript. FJAN designed the study, helped with data analyses, and gave feedback on the manuscript. All authors gave final approval for publication.

Figure captions

Figure 1. Map showing the location of the stations sampled during September 20–23, 2018. At each station triplicate sediment cores were collected and the top 0–2 cm sediment surface sliced. The study area is located in the Gulf of Finland (Baltic Sea) nearby the Tvärminne Zoological Station (TZS). The numbers denote each station name. Stations 11, 12, 15, 16 were grouped as "Low OC", and stations 7, 10, 13 as "High OC" based on the % OC content. The map layer is © OpenStreetMap contributors.

Figure 2. The boxplots show the species richness Chao1 and Shannon's H alpha diversity index for the three taxonomic groups studied in the sediment surface in the Low OC and High OC stations. The data are based on 18S rRNA sequences extracted from the RNA-seq data, with (a–c) showing Chao1 and (d–f) showing Shannon's H. Note the different scale on the y-axes between the three taxonomic groups. The P values show the results from One-Way ANOVA tests between the Low OC and High OC (only shown if statistically significant). The outliers denote, circles: 1.5–3 box lengths from the median, and stars: 3 or more box lengths from the median.

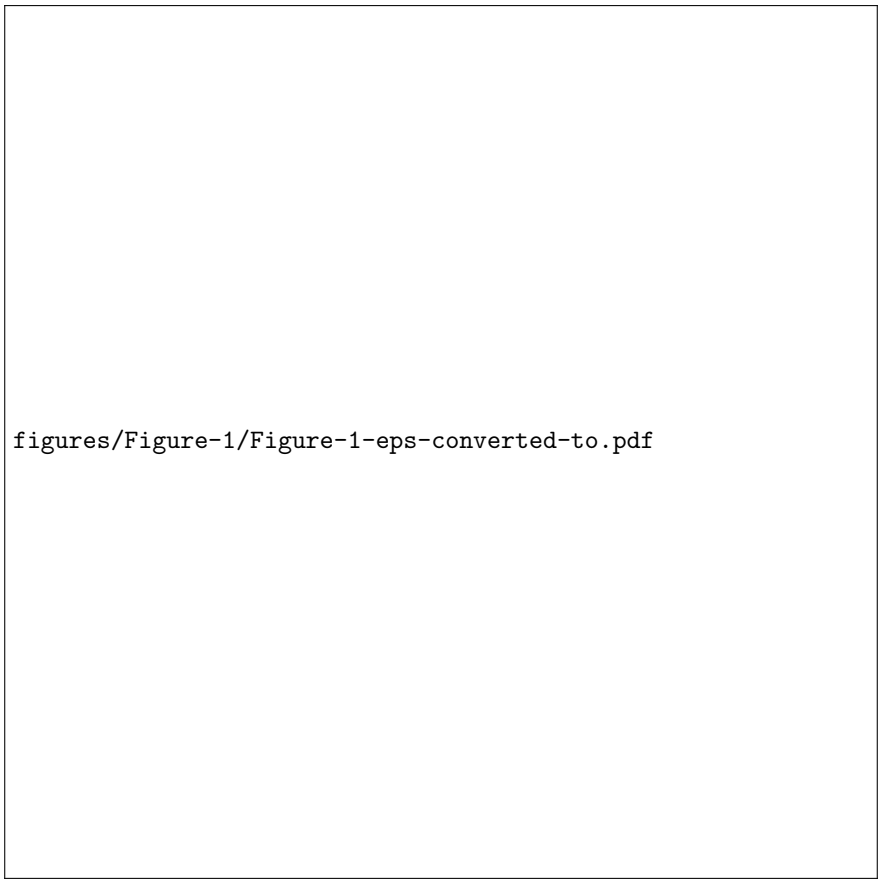
Figure 3. NMDS plots showing the beta diversity of the three studied taxonomic groups in the sediment surface, featuring (a)*Nematoda* , (b), *Foraminifera* , and (c) *Ciliophora* . The beta diversity was based on the 18S rRNA data and the Sørensen index (presence/absence, labels show station numbers). The light blue shaded areas denote Low OC stations, while dark blue shaded areas denote High OC stations. The P values show the results from PERMANOVA tests between the Low OC and High OC stations.

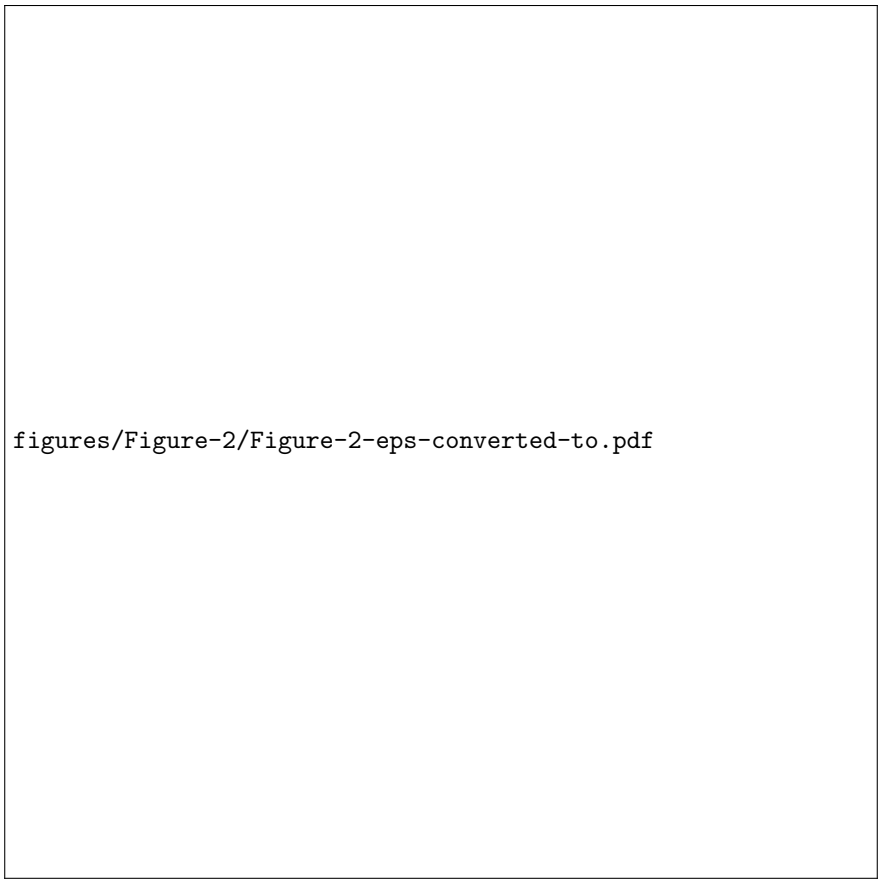
Figure 4. CCAs showing the distribution of (a) *Nematoda* , (b) *Foraminifera* , and (c) *Ciliophora* among the Low OC stations (light blue circles) and High OC stations (dark blue circles). The data was based on the relative abundances o genera for each taxonomic group. The grey triplots shows the direction of to the measured abiotic variables (water depth, sediment % OC and % N, plus pore water NH_4^+ and PO_4^{3-}) in

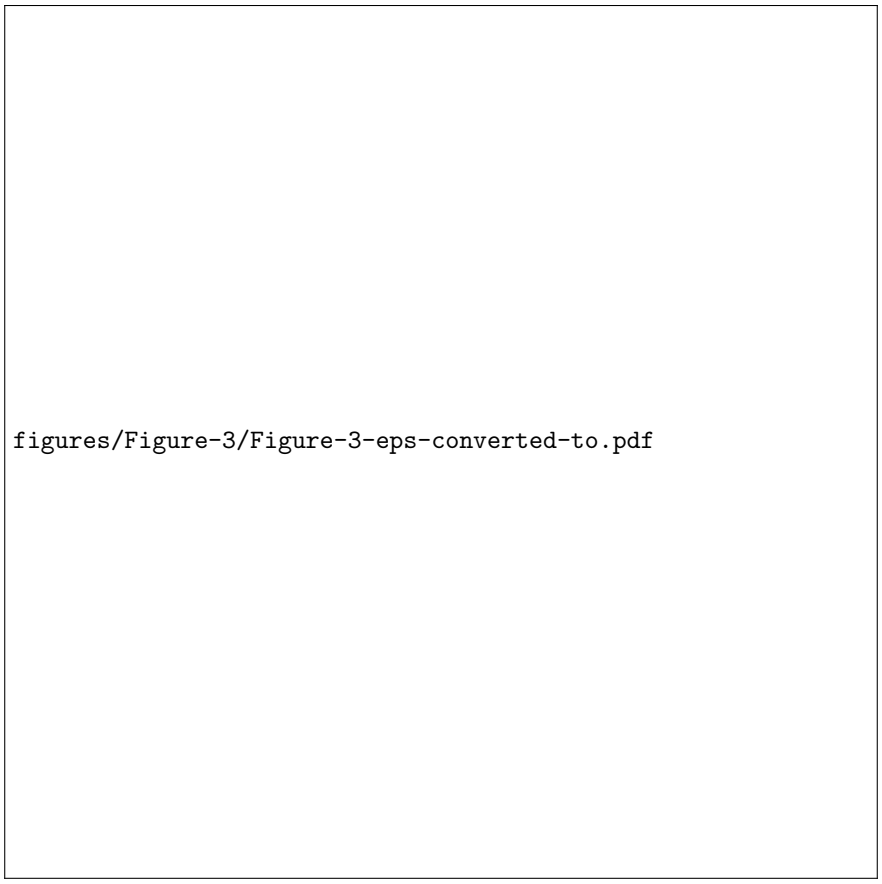
relation to the community composition. Each circle represents one sediment core. The *P*values shows the statistical significance (PERMANOVA) between the abiotic data and community composition when tested between Low OC and High OC stations.

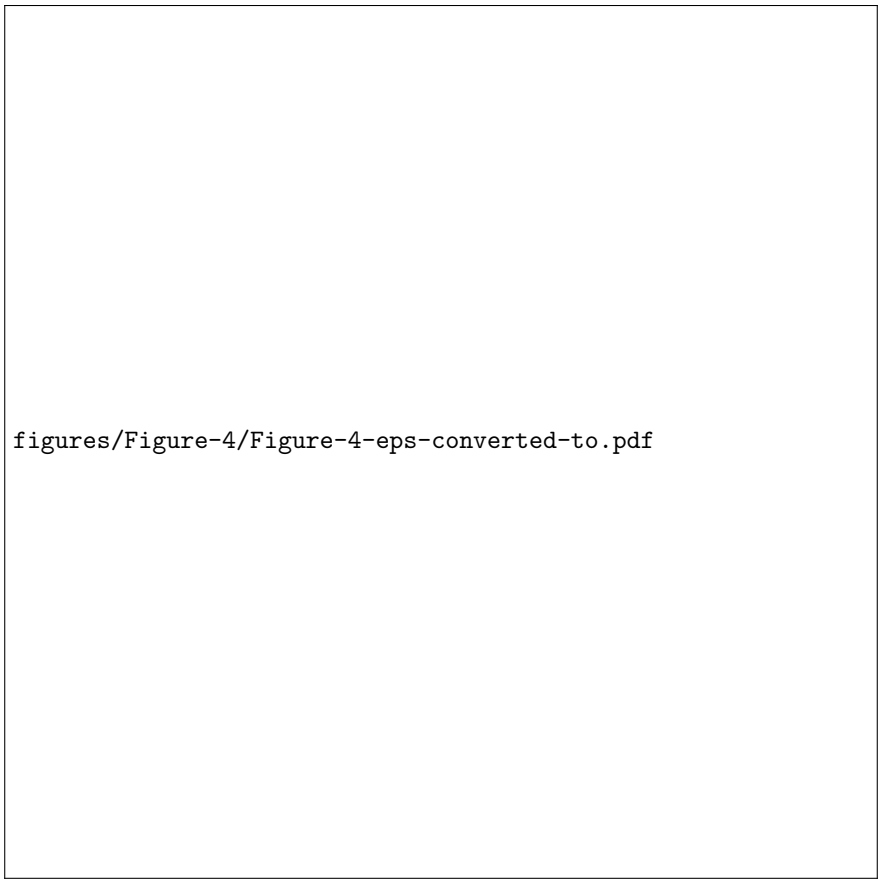
Figure 5. The stacked bars show the taxonomic classifications for the groups (a) *Nematoda* , (b) *Foraminifera* , and (c)*Ciliophora* based on 18S rRNA data (nematodes and forams classified against NCBI NT, and ciliates against the SILVA database). The y-axis shows the station names, their water depth (m), and replicates denoted with A, B, C. The x-axes show the relative abundance (%) within each taxonomic group. Taxonomic classifications that are significantly different between Low OC and High OC sites mentioned in the results have been indicated with bold text.

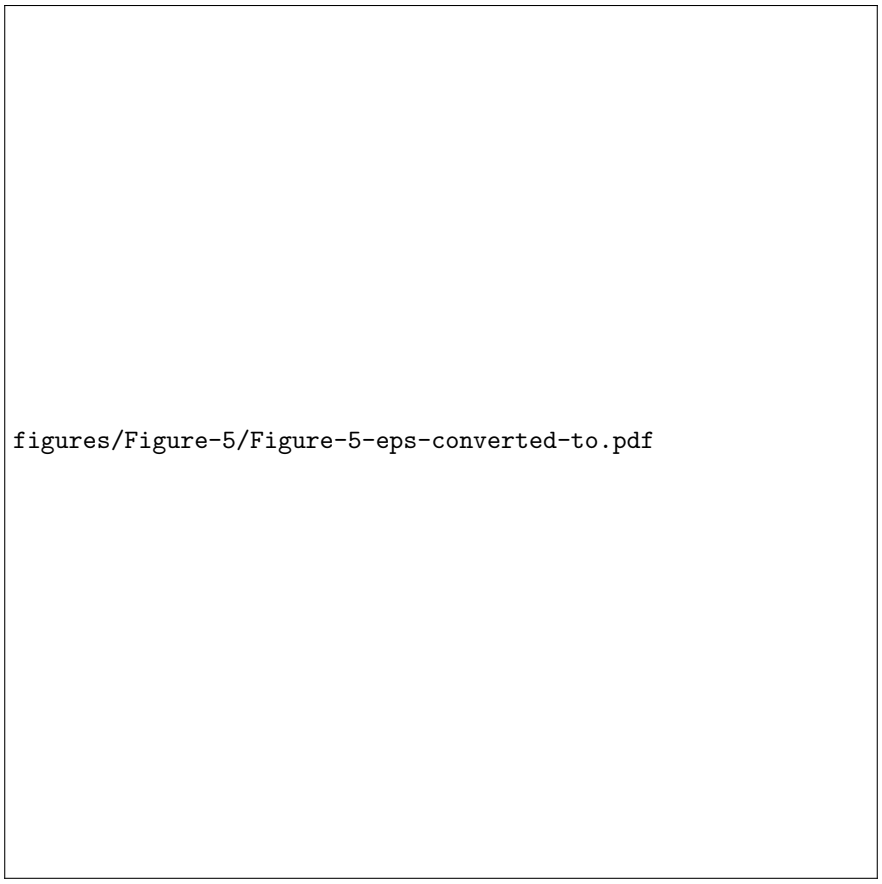
Figure 6. *Nematoda* genera were classified into a feeding type category according to Wieser (1953) and the plot is based on the sum of all classifications between the Low OC and High OC stations. DESeq2 statistical analyzing showed significant differences for all feeding types ($FDR < 0.05 = *$, $FDR < 0.01 = **$). Negative log2 fold change values indicate a higher prevalence at the High OC stations (dark blue circles), while positive values indicate a higher prevalence at Low OC (light blue circles). The errors bars show the standard error.

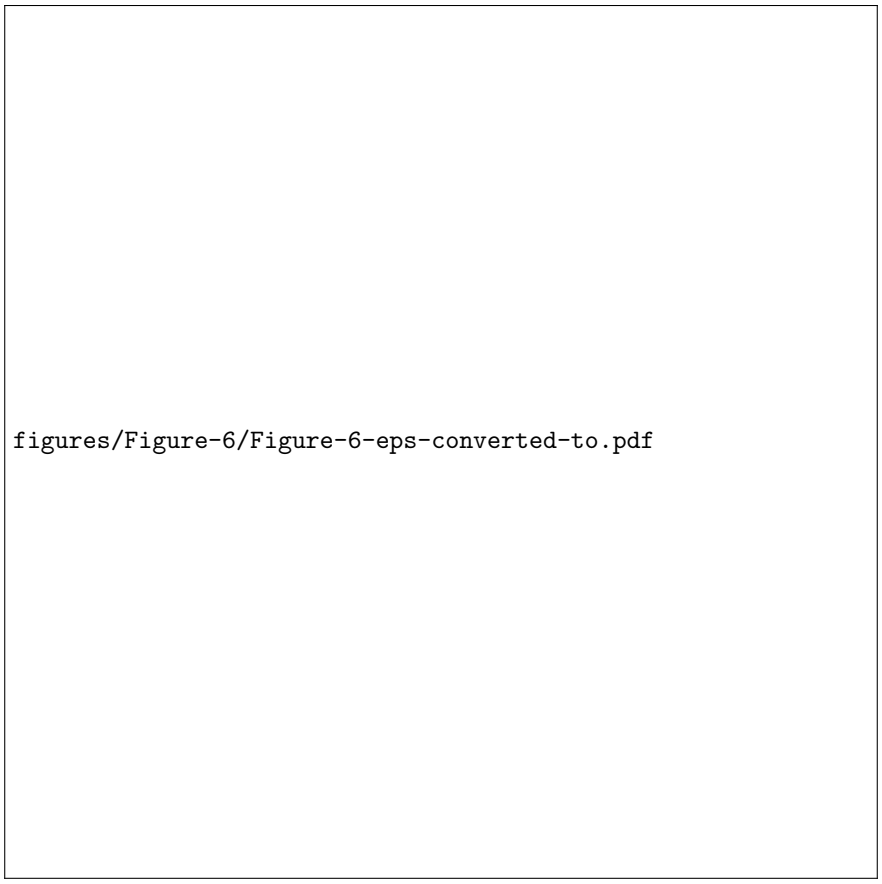












Hosted file

Table 1.docx available at https://authorea.com/users/295567/articles/424201-high-throughput-shotgun-sequencing-of-erna-reveals-taxonomic-and-derived-functional-shifts-across-a-benthic-productivity-gradient