

# SCIENTIFIC REPORTS



OPEN

## Sample size effects on the assessment of eukaryotic diversity and community structure in aquatic sediments using high-throughput sequencing

Francisco J. A. Nascimento<sup>1,4</sup> , Delphine Lallias<sup>2</sup>, Holly M. Bik<sup>3</sup>  & Simon Creer<sup>4</sup> 

Understanding how biodiversity changes in time and space is vital to assess the effects of environmental change on benthic ecosystems. Due to the limitations of morphological methods, there has been a rapid expansion in the application of high-throughput sequencing methods to study benthic eukaryotic communities. However, the effect of sample size and small-scale spatial variation on the assessment of benthic eukaryotic diversity is still not well understood. Here, we investigate the effect of different sample volumes in the genetic assessment of benthic metazoan and non-metazoan eukaryotic community composition. Accordingly, DNA was extracted from five different cumulative sediment volumes comprising 100% of the top 2 cm of five benthic sampling cores, and used as template for Illumina MiSeq sequencing of 18 S rRNA amplicons. Sample volumes strongly impacted diversity metrics for both metazoans and non-metazoan eukaryotes. Beta-diversity of treatments using smaller sample volumes was significantly different from the beta-diversity of the 100% sampled area. Overall our findings indicate that sample volumes of 0.2 g (1% of the sampled area) are insufficient to account for spatial heterogeneity at small spatial scales, and that relatively large percentages of sediment core samples are needed for obtaining robust diversity measurement of both metazoan and non-metazoan eukaryotes.

Predicting how benthic diversity and community structure respond to environmental change will continue to be an important challenge for aquatic ecology. Such a task is dependent on fast and reliable methods to assess diversity, particularly in ecosystems like soft-sediment habitats, where most eukaryotic diversity is comprised of microscopic communities<sup>1,2</sup> crucial to a number of important ecosystem processes<sup>3-6</sup>. Despite their demonstrable contributions to important ecological functions, the levels of eukaryotic and, in particular, metazoan diversity are not well known, due to a number of practical limitations involved in the taxonomic and ecological investigations of microscopic eukaryotic communities<sup>7</sup>. Among these, the level of time and financial resources necessary to cover such broad taxonomic scales are perhaps the most important<sup>1,7,8</sup>.

The application of high-throughput sequencing (HTS) to the study of benthic eukaryotes has the potential to significantly advance our knowledge of interstitial eukaryotic diversity<sup>8-10</sup>. HTS techniques can be applied for species identification from samples where organisms have been separated and isolated from the sediment/soil matrix or water before the analysis<sup>10</sup>, or from environmental DNA (eDNA). Environmental DNA is here defined as the genetic material attained from samples (soil, sediment, water) irrespective of whether the DNA is intracellular or extracellular, or associated with living or dead biomass<sup>11</sup>.

Establishing an appropriate and standardized sampling strategy is critical to the use of HTS for studies of benthic eukaryotic biodiversity, particularly since the application of HTS techniques to benthic ecology questions is

<sup>1</sup>Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm, Sweden. <sup>2</sup>GABI, INRA, AgroParisTech, Université Paris Saclay, 78350 Jouy-en-Josas, France. <sup>3</sup>Department of Nematology, University of California Riverside, Riverside, CA, USA. <sup>4</sup>Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences, Bangor University, LL57 2UW, Bangor, United Kingdom. Correspondence and requests for materials should be addressed to F.J.A.N. (email: [francisco.nascimento@su.se](mailto:francisco.nascimento@su.se))

relatively recent<sup>10,12,13</sup>. Sediment ecosystems often include a matrix of different particles and aggregates that form highly complex and heterogeneous physical environments at both local (e.g. biogenic structures like feeding pits or fecal mounds create heterogeneity at a centimeter scale) and at regional spatial scales where benthic habitats encompass different habitats such as sandy or soft-sediments kilometres apart<sup>14,15</sup>. At finer scales, heterogeneity and spatial structure can hinder the detection of ecological patterns and processes at ecosystem scales<sup>15,16</sup>. In particular, the large number of niches contained within small spaces in sediments<sup>14</sup> leads to a heterogeneous distribution of eukaryotic microbial communities<sup>16,17</sup>. In addition, DNA from aquatic organisms is itself often distributed heterogeneously<sup>18</sup> and in benthic habitats it can adsorb more or less efficiently to different minerals<sup>19</sup>, creating additional spatial heterogeneity. This DNA adsorption capacity can differ greatly between sediment types, with clay minerals typically having a hundred-times higher adsorption capacity than sandy sediments<sup>20</sup>. The study of biodiversity using DNA sampling is therefore dependent on several factors like the concentration of DNA in the sampled habitat, DNA capture and extraction efficiency and sample interference (e.g. inhibition)<sup>21</sup>.

Common strategies to deal with small scale heterogeneity include the collection of large samples (often unpractical and costly) or several smaller samples that are pooled, homogenized and from which a smaller sub-sample is taken and analysed, that is considered to be representative of the entire mix<sup>17</sup>. Lanzén *et al.*<sup>22</sup> showed that pooling increasing numbers of DNA extraction replicates significantly improved eukaryotic diversity estimates for both protists and metazoans in marine sediments. Conversely, recent works using HTS for environmental monitoring of ciliates indicate that replicates from small sediment samples are similar to each other<sup>23</sup>. The strategy of extracting DNA directly from small homogenised samples of sediment without subsequent isolation of organisms from the sediment matrix is often utilized in works employing HTS for the assessment of benthic microeukaryotic diversity<sup>24–26</sup>, although such studies often emphasize diversity patterns in single-celled eukaryotes (e.g. protists) (See Supplementary Table S1 for examples of previous work that assessed eukaryotic diversity in aquatic sediments using high-throughput sequencing)<sup>27–36</sup>. Additionally, a number of works have successfully employed eRNA in environmental monitoring assessments of marine eukaryotes<sup>23,37–40</sup>. Even though the sampling of eRNA presents additional storage and preservation challenges, it can depict accurately the active fraction of interstitial communities<sup>37,38,41</sup>. Nevertheless, eDNA has a higher persistence than eRNA and has been shown to capture more representative measures of interstitial biodiversity than eRNA<sup>41</sup>.

However, how much such strategies can tell us about sediment metazoan diversity is not clear as small samples do not seem to be effective in capturing metazoan meiofauna diversity. Meiofaunal organisms are physically larger (38 µm–1 mm in size), have patchy distributions in space and time, and most likely have lower effective population sizes compared to bacterial and protist populations<sup>14</sup>. As such, when targeting meiofauna diversity, isolation from larger sediment samples is recommended<sup>12</sup>. However, meiofauna studies typically employ decantation or density extraction and sieving in order to concentrate metazoan organisms and physically separate them from sediment grains. Ecologically important organisms such as single-celled microeukaryotes and taxa < 38 µm in size are lost during these physical extraction/isolation protocols. There are obvious advantages in the use of the same sample for HTS diversity assessments of multiple communities of different size fractions (eg. protists, bacteria and meiofauna)<sup>39</sup>. Only by measuring these metrics in the same sample will we be able to capture and understand the interspecific interactions occurring in small spatial scales and how they affect sediment diversity.

As such, one important methodological question that needs to be addressed for benthic HTS is what sample volume from a homogenized mixture can adequately represent the diversity and community structure of both microeukaryotes and metazoans in the sediment. Sample size can affect the measurement and interpretation of biodiversity metrics, particularly in species rich ecosystems<sup>42</sup>. Insufficient sample sizes are responsible for non-trivial biases on estimations of beta-diversity that are hard to predict<sup>42</sup>. Since diversity metrics are essential tools for benthic ecological research, with the emergence of HTS, it is vital to understand how sample volumes influences biodiversity metrics in aquatic sediments.

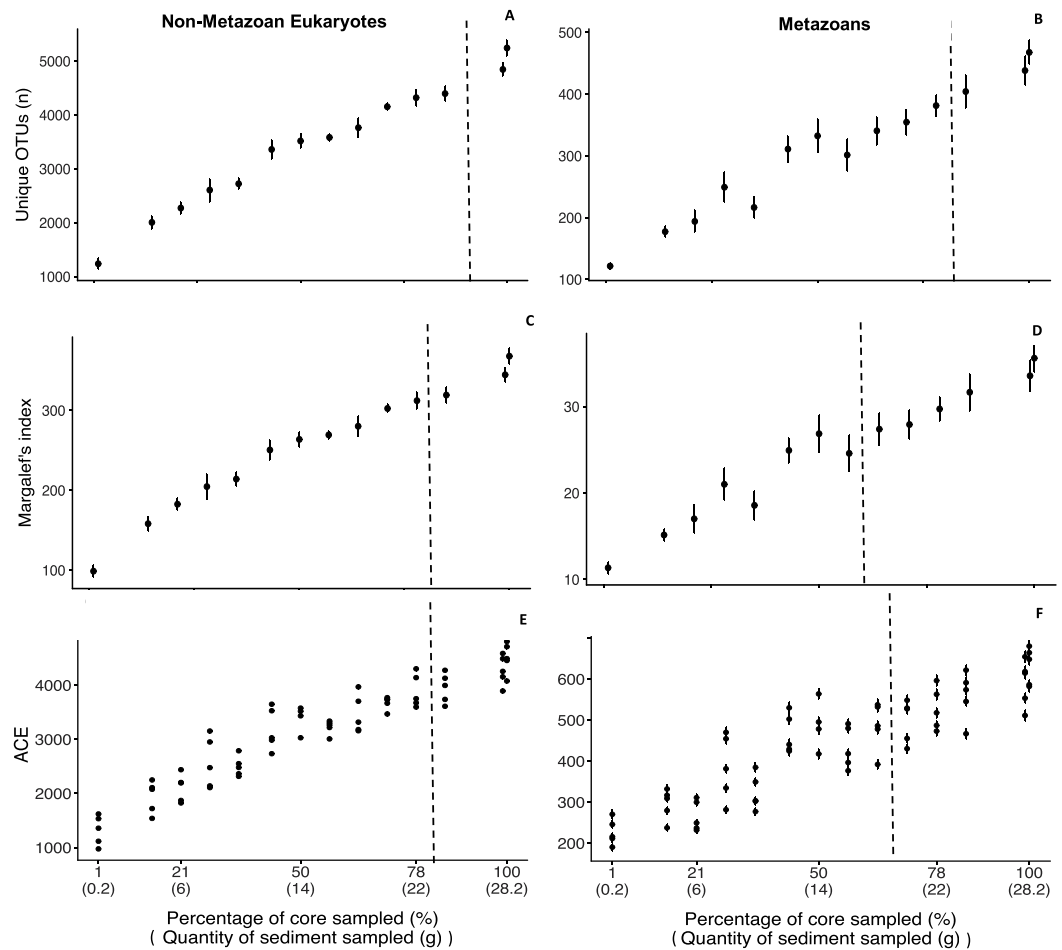
Soil microbial ecologists have found sample volume to have an effect on the patterns of soil microbial biomass<sup>43</sup> and on how accurately microbial community structure and diversity is assessed<sup>44–46</sup>. For example, soil volumes of 10 g were related to higher richness, diversity and evenness<sup>46</sup> and with lower variability between replicates than smaller sample volumes<sup>44</sup> (see Supplementary Table S2 for an overview of studies that investigated effects of sample volumes on diversity assessments). Nevertheless, few studies have looked at the effects of sample volumes using HTS in aquatic sediments.

Therefore, we address in our study an important question in biodiversity assessments of marine sediments using HTS: what percentage of a core sampling unit is a representative and appropriate sample volume for a reliable measurement of both micro-eukaryotic and metazoan diversity and associated community structure?

Consequently, we devised a novel experimental design that analysed metazoan and non-metazoan eukaryotic diversity and community composition in replicated samples of five different sediment volumes (0.2, 4, 6, 8 and 10 g) that together comprised 100% surface area of a sampling core. Sampling cores are generally considered the best quantitative sampling technique for benthic communities in marine clay and silt sediments<sup>47</sup> and here we employed a sampling core commonly used in both field and experimental studies of benthic ecology<sup>3–5</sup>. Following environmental DNA extractions, we sequenced the V4 region of the 18 S nuclear small subunit (nSSU) ribosomal RNA (rRNA) gene from these different sediment quantities and used ca. 400 base pair amplicons to assess the impact of sample volume on alpha and beta-diversity measurements. Our findings contribute to the development of consistent protocols to facilitate the comparison among ecological studies on benthic eukaryotic communities.

## Results

**HTS data output.** Community analysis was performed on both the 96% and the 99% OTUs and produced similar results. For conciseness, we present the 96% OTU data here and the 99% OTU data in Supplementary Information (See Supplement Table S3 for statistics regarding the 99% OTUs).



**Figure 1.** Diversity measurements for Non-Metazoan Eukaryotes (A,C,E) and Metazoans (B,D,F) in the different treatments. Figure panels show number of unique OTUs (A,B), Margalef Diversity index (C,D) and ACE index (E,F). Dotted line indicates the cutoff point after which the differences to the 100% treatment stop being significant (see Table 1). Diversity measurements were performed on transformed data. Data represents 96% OTUs.

In total the Illumina Miseq sequencing of eukaryotic 18S rRNA amplicons yielded 10 320 000 raw paired-end reads from 25 samples from which 7 776 371 quality-filtered reads were assembled. After chimera removal there was an average of 305 997 sequences per sample (min-168 672; max- 572 017; Supplement Table S4). Clustering at 96% OTU similarity produced 10 220 OTUs assigned to domain Eukaryota after singleton and doubleton removal, of which 4870 were non-metazoan eukaryotes, 870 OTUs were from metazoan taxonomic groups and 4480 could not be assigned to Order. OTU accumulation plots for Non-metazoan Eukaryotes and Metazoans are presented in Supplementary Information (Supplementary Fig. S1).

**Alpha diversity differences with sample volume.** The number of unique OTUs was lowest in the 1% of the total sampled area (0.2 g) with an average of 1238 and 121 OTUs for non-metazoan eukaryotes and metazoans, respectively, and increased with sample volume to a maximum of 5243 and 468 eukaryotic and metazoan OTUs in the treatment with 100% of the total sample area (28.2 g) (Fig. 1A,B). The difference in the number of unique non-metazoan OTUs between the 100% treatment and all other treatments was significant for all sample volumes below 85% of the total area sampled (24 g), but not for any remaining sample volumes above that value (Table 1). A similar pattern was seen for the number of unique metazoan OTUs but with a lower threshold, at 64% of the total area sampled (18 g) (Table 1). There was also a significant effect of sample volume on diversity for both non-metazoan eukaryotes and metazoans (Fig. 1C–E, Table 1). Pairwise comparisons between the treatment comprising 100% of the sample area (28.2 g) and all other treatments show that the differences in Margalef and ACE for non-metazoan eukaryotes were significant for all sample volumes below 71% (20 g) and 78% (22 g), respectively, but not for larger sample volumes (Table 1). For metazoans that cut-off was at 57% of the total area sampled (16 g) for Margalef and 63% (18 g) for ACE index (Table 1), respectively.

**Beta-diversity differences with sample volume.** Figure 2 shows a simplified NMDS ordination of the non-metazoan eukaryotic (Fig. 2A) and metazoan (Fig. 2B) community composition. An NMDS and PCoA plot with all treatments for the 96% OTUs can be seen in the Supplementary Information (Supplement Fig. S2-A,S2-B,

	Community	ANOVA F statistics	ANOVA p value ANOVA	Lowest % of core sampled significantly different from the 100% treatment
Percentage of metazoan taxa	NA	4.1	<0.001	1% (0.2 g)
Number of unique OTUs	Non-Metazoan Eukaryotes	67	<0.001	85% (24 g)
	Metazoans	23	<0.001	78% (22 g)
Margalef index	Non-Metazoan Eukaryotes	57	<0.001	71% (20 g)
	Metazoans	17	<0.001	57% (16 g)
ACE richness	Non-Metazoan Eukaryotes	51	<0.001	78% (22 g)
	Metazoans	19	<0.001	63% (18 g)
		PERMANOVA F Statistics	PERMANOVA p value	Lowest % of core sampled significantly different from the 100% treatment
Beta-Diversity	Non-Metazoan Eukaryotes	2.5	<0.001	43% (12 g)
	Metazoans	1.9	<0.001	71% (20 g)

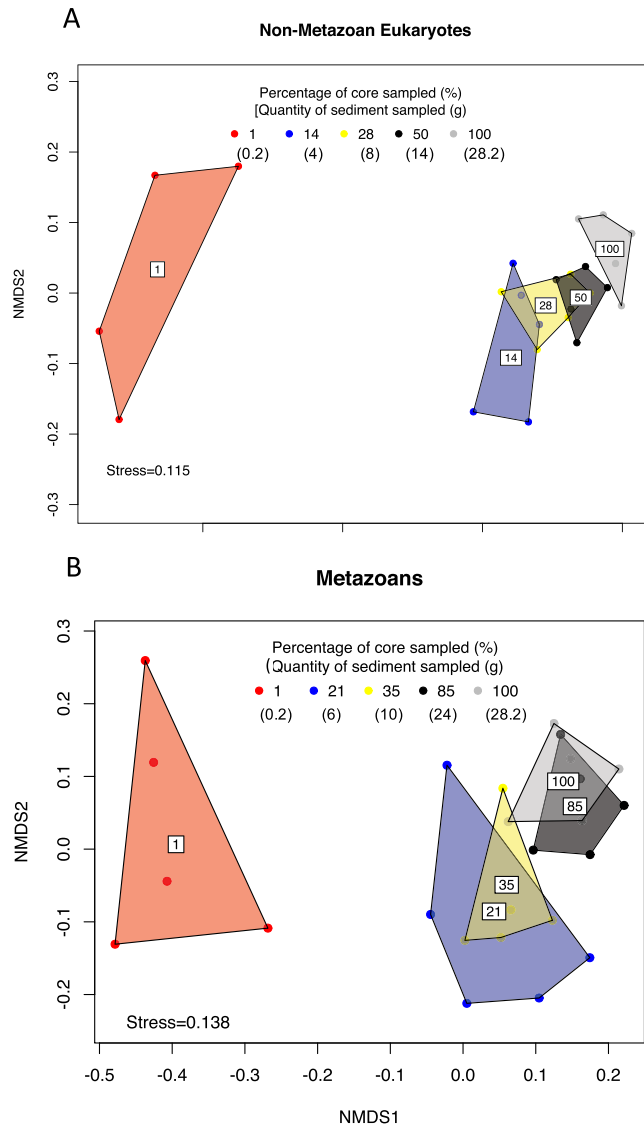
**Table 1.** Summary of statistical tests comparing the different treatments with the treatment comprising 100% of the sampled area (96% OTUs). One-way parametric analysis of variance ANOVA was used to test for the effect of sample volume. P values smaller than 0.01 indicate a significant effect of sample volume. Pairwise comparisons between the 100% treatment and the other 13 treatments were performed with the Tukey HSD test. Effects of sample volumes on beta-diversity differences among treatments were tested with PERMANOVA. Differences between the 100% treatment and the other 13 treatments for beta-diversity was tested with PERMDISP. The treatment with the highest sample volume still different from the treatment comprising 100% of the sampled area in each of the variable is presented in the last column of the table.

for non-metazoan eukaryotes and metazoan NMDS, and Supplement Fig. S3-A,S3-B for PCoA ordination, respectively). PERMANOVA analysis showed a significant effect of sample volume in both non-metazoan eukaryotic and metazoan community beta-diversity (Table 1). PERMDISP revealed significant overall differences among treatments. Non-metazoan community composition of the 100% treatment (28.2 g) was different from all sample volumes below 43% of the total sampled area (12 g) but not significantly different from any of the sample volumes equal or larger than this value (Table 1). The differences in beta-diversity between the treatment comprising 100% of the area sampled (28.2 g) and all other sample volumes are not significant for the 71% treatment (20 g) and above. These dissimilarities among treatments in relation to the 100% of the area sampled decreased clearly with increased sample volume (Fig. 3). Data analysis with 99% OTUs presented similar results (see Supplement Table S3 for statistics, Supplement Fig. S4 for simplified NMDS and Fig. S5 for PCoA).

Unsurprisingly, differences between the 100% treatment and smaller sample volumes (0.2 g, 4 g, 8 g, 10 g and 12 g) were mostly driven by nestedness, while turnover was more influential in explaining overall beta-diversity differences between larger sample volumes and the 100% treatment. This pattern was seen for both Non-Metazoan Eukaryotes and for Metazoans (Supplement Fig. S6A,B, respectively).

**Additional sample volumes and read combination.** Combining reads of different sample volumes had no effect on beta-diversity measurements. As it is possible to see in both the NMDS plot and in the weighted PCoA (Fig. 4), there is considerable overlap between the two ways of analysing sample volume of 10 g here tested (6 g + 4 g vs 10 g) for both metazoan and non-metazoan eukaryotic beta-diversity. We found no statistical differences in beta-diversity between these two ways of obtaining 10 g of sediment (PERMANOVA,  $p = 0.24$  and  $p = 0.65$  for metazoans and non-metazoan eukaryotes, respectively). In fact, Sørensen's dissimilarity among the 5 replicates of the 10 g treatment is higher than the average dissimilarity between the pairs of 6 g + 4 g and 10 g for each replicate (Fig. 5). Both these two analyses suggest that we obtain very similar representations of beta-diversity if we pool 6 g + 4 g or if we analyse 10 g directly. Additionally, we performed a similar test to investigate whether there were significant differences in beta-diversity if we pooled 8 + 6 g vs 10 + 4 g to get 14 g and 8 g + 6 g + 4 g vs 10 g + 8 g to get 18 g. We found the same pattern as mentioned above for 10 g vs 6 + 4 g (Supplement Fig. S7). No statistical differences were found in beta-diversity (PERMANOVA,  $p = 0.53$  and  $p = 0.92$  for 14 g and 18 g, respectively) and higher dissimilarity between the replicates of the same pooling combination than the replicates obtained with different pooling combinations (Supplement Fig. S8). Taken together, these results show clearly that there is no effect of pooling samples and that the differences seen in our study are indeed due to sample volume.

Our data analysis was performed with relative OTU abundances, suggested by McMurdie & Holmes<sup>48</sup> as an appropriate and effective alternative to rarefying. However, the same authors warn that artefacts due to varying sampling depths can persist even when using relative OTU abundance<sup>48</sup>. Since we cumulated reads numbers from 0.2–10 g volumes to represent up to 100% of the cores, it is possible that there was an over-inflation of rare taxa in the larger sample volumes that would affect our diversity metrics. However, as discussed above and shown in Figs 4 and 5 and Figs S7 and S8 in Supplementary Information, combining reads of different sample volumes had no significant effects on beta-diversity assessment. Furthermore, an analysis with a dataset trimmed to contain only the 30 most abundant non-metazoan eukaryotic taxa in each sample volume, produced similar results to the whole data set in terms of the effect of sample size on beta-diversity measurements (PERMANOVA,  $F = 8.0$ ,  $p = 0.001$ , NMDS plots presented in Fig. S11 Supplementary Information). Differences in beta-diversity between the 100% treatment and the six smaller samples sizes (0.2 g, 4 g, 6 g, 8 g, 10 g and 12 g) tested with PERMDISP were all significant, but not for the remaining sample sizes above 12 g.

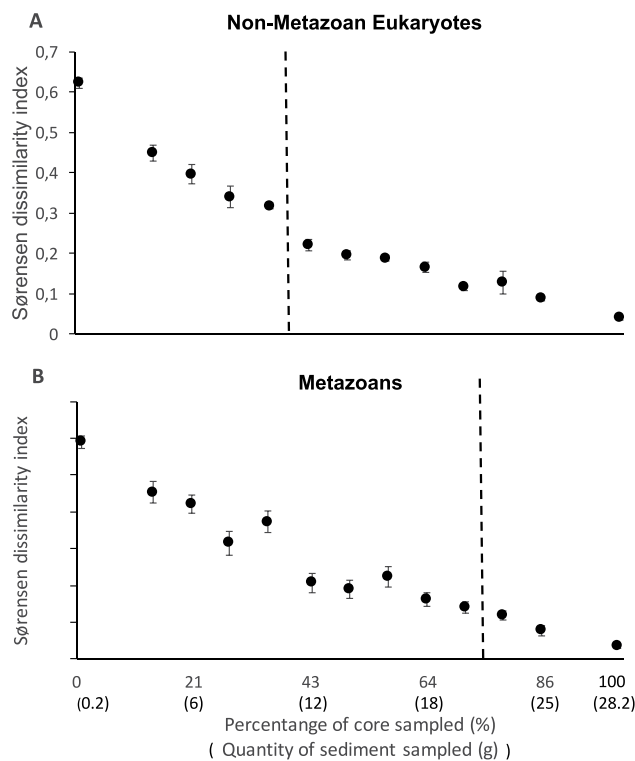


**Figure 2.** Simplified non-metric multidimensional scaling (NMDS) analysis of the Sørensen beta-diversity matrix based on transformed OTU table for Non-Metazoan Eukaryotes (A) and Metazoans (B). Differences between the different treatments and the 100% treatment stopped being significant at 43% of sediment analyzed for Non-Metazoan Eukaryotes and in the 78% treatment for Metazoans (PERMANOVA, PERMDISP, see Table 1). For better visualization, only the treatment comprising 100% of the sampled volume, 3 treatments significantly different from the 100% treatment (1, 14, 28% for Non-Metazoan Eukaryotes and 1, 21 and 35% for Metazoans), and one treatment not significantly different from the 100% treatment (50% for Non-Metazoan Eukaryotes and 85% for Metazoans) are here shown. Data represents 96% OTUs.

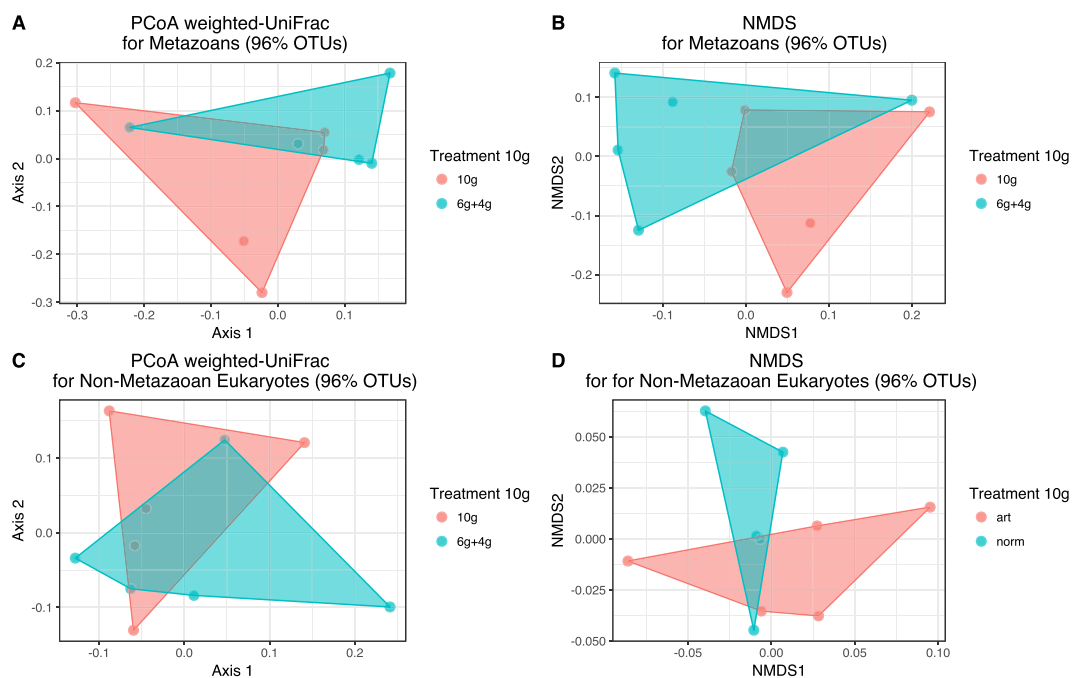
**Taxonomic composition.** The percentage of OTUs belonging to metazoan groups was different among sample volumes (Fig. 6), with the smallest sample volume (1% of the total sampled area, corresponding to 0.2 g) containing significantly less metazoan OTUs than all the other treatments (ANOVA,  $p < 0.001$ , Table 1). No additional significant differences among other sample volumes in metazoan vs non-metazoan eukaryote taxa were detected. Generally, the most abundant non-metazoan eukaryotic groups were the dinoflagellates, followed by diatoms and ciliates (Fig. 7A and Supplement Fig. S9-A). Copepod OTUs dominated the metazoan dataset comprising more than 70% of the relative abundance. Bivalvia (mostly the clam *Limecola balthica*) and nematodes OTUs were the other most frequent metazoan groups in our study (Fig. 7B and Supplement Fig. S9-B).

## Discussion

**Effects of sample volume on both alpha and beta-diversity.** Our results show clearly that sample volume affected all non-metazoan eukaryotic and metazoan diversity metrics investigated here. Sample volume had a significant effect on percentage of metazoan OTUs, number of unique OTUs, Margalef and ACE diversity indexes both for non-metazoan eukaryotes and metazoans (Fig. 1, Table 1).

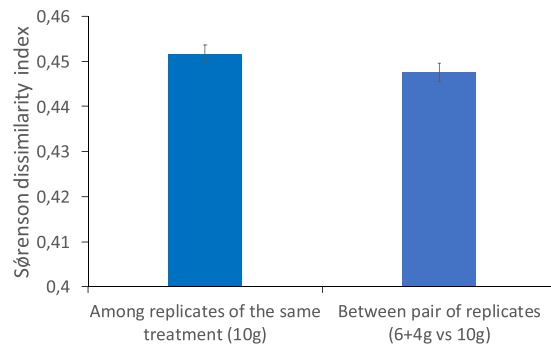


**Figure 3.** Average Sørensen dissimilarity index in each treatment in relation to the treatment where 100% of the sampled area was analyzed ((A)-Non-Metazoan Eukaryotes; (B)- Metazoans). Circles represent average dissimilarity in relation to the 100% treatment and error bars SE. Dotted line indicates the cutoff point after which the differences to the 100% treatment stop being significant (see Table 1). Sørensen dissimilarity index was calculated using transformed OTU data. Data represents 96% OTUS.

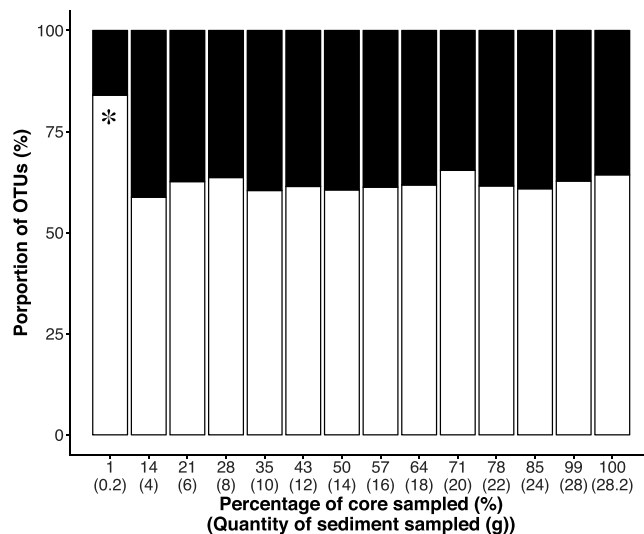


**Figure 4.** PCoA (A,C) and NMDS (B,D) ordinations comparing two different strategies for obtaining 10g with the sample volumes used in this study: 6 + 4g (in blue) vs 10g (in red). No statistical differences were seen between the two treatments (PERMANOVA,  $p = 0.24$  and  $p = 0.65$ ). Data represents 96% OTUS.





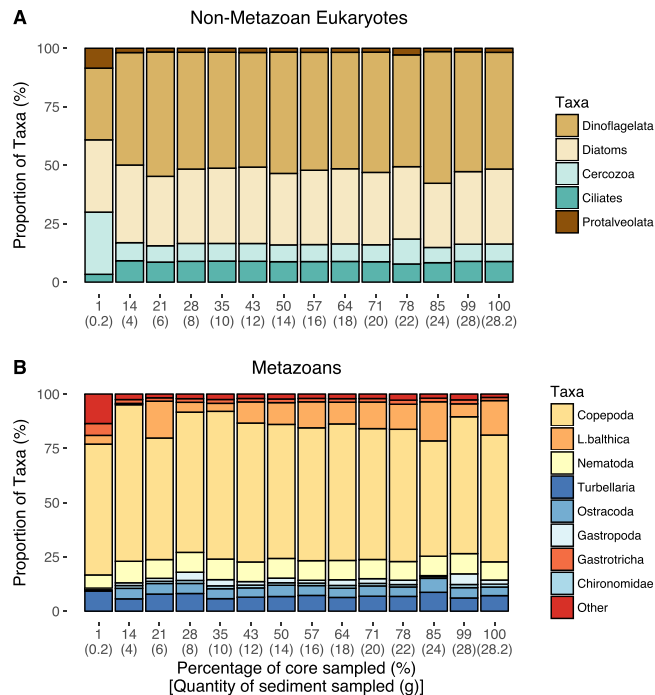
**Figure 5.** Average Sørensen dissimilarity index among replicates of the 10 g and distance between 10 g vs 6 g + 4 g of the same replicate (based on data for Metazoans). Data represents 96% OTUs.



**Figure 6.** Differences among treatments in proportion of Metazoan (full bar) vs Non-Metazoan Eukaryotes (empty bars). Asterisk indicates a significant difference among treatments in proportion of metazoan OTUs. These proportions represent transformed OTU data, from 96% OTUs.

These results are in accordance with previous studies that found sample volume to have a significant influence when investigating the diversity and community composition of soil bacteria<sup>17,44</sup>. In addition, Penton *et al.*<sup>46</sup> using HTS to look at soil fungal and bacterial communities found that alpha diversity, richness and evenness to be larger with larger sample volumes showing a similar pattern to our results. However, related effects of sample volume on richness and alpha diversity were not seen in marine meiofauna in Brannock and Hananych<sup>12</sup>. Although these authors compared alpha diversity metrics among samples of different volumes, only two replicates were employed, which might account for the different results. Brannock & Hananych<sup>12</sup> also used a different set of primers on fully marine sediments that generally include more diverse communities than soft-sediment brackish habitats like the Baltic Sea<sup>14</sup>.

Sample volume was also a significant factor when analysing beta-diversity for both non-metazoan eukaryotes and metazoans, with the smallest sample volume clearly separately from all other groups (Fig. 2). Our results indicate that dissimilarity among treatments in relation to the 100% treatment decreases with sample volume (Fig. 3), showing that sample volume is an important factor when assessing sediment beta-diversity. Indeed, the larger the sample volume, the more animals are captured and the more similar the samples are to the 100% treatment. Sample volume effects on beta-diversity and community composition were seen on different taxonomic groups from other studies, including bacteria, eukaryote and metazoans in both terrestrial and aquatic ecosystems<sup>12,17,44,46</sup>. While the effects of sample volume on beta-diversity were significant in our metazoan and non-metazoan eukaryotes datasets (Fig. 2, Table 1), there were important differences between these two taxonomic groups. Our results suggest that the analysis of a much larger percentage of the core is needed for metazoans (78%) than for non-metazoan eukaryotes (43%) in order to achieve a representative picture of the real beta-diversity, indicating a higher degree of patchiness for the metazoan community. Such small spatial scale patchiness has also been reported for small benthic metazoans like meiofauna<sup>12,14</sup> as these animals tend to cluster in microsites rich in organic matter content<sup>14,49</sup>. A common strategy in studies focusing on meiofauna community structure is to isolate individuals from large sediment sample sizes by decantation or density extraction in a



**Figure 7.** Proportion of the most abundant taxonomic groups in each treatment: (A) Top five most abundant Non-Metazoan Eukaryotes in each treatment; (B) Top nine most abundant Metazoan taxa in each treatment. Data represents 96% OTUs.

38–45  $\mu\text{m}$  sieve before DNA extraction. Decantation concentrates living individuals from a large sample size into a smaller volume and reduces the quantity of extracellular DNA, DNA from dead/decaying, resting stages and the quantity of PCR-inhibiting humic acids in the analysed sample. Such a strategy is recommended by previous studies for HTS assessments of meiofauna diversity<sup>12</sup>. However, decantation will necessarily involve the loss of information from protists and eukaryotes smaller than 38  $\mu\text{m}$  and soft-bodied animals in the meiofauna fraction. For studies aiming at assessing biodiversity of both the larger meiofauna and microeukaryotes, the effects of decantation must therefore be considered. Combining multiple sub-samples for microbial eukaryotes, with sieving and isolation of the larger meiofauna from the remaining sediment sample, are a feasible alternative strategy when aiming for simultaneous recovery of benthic taxa representing different size fractions and trophic levels<sup>50</sup>. The extracted DNA for each size fraction could then be PCR amplified and bioinformatically pooled after sequencing for effective diversity assessments. In addition, this strategy is useful in studies using sampling equipment with larger surface areas such as Van Veen grab sampler. This sampling equipment has a total surface area considerably larger than the core used in our study and is often used in environmental monitoring studies that employ HTS in sediment surveys of diversity (e.g.<sup>23,40,41,51–53</sup>). As such, sampling volumes equivalent to for example 43% and 78% of the surface area of a Van Veen grab sampler will not be realistic for logistical reasons and strategies that employ combination of multiple sub-samples for microeukaryotes together with animal isolation for metazoans are appropriate<sup>23,39–41,51–53</sup>. Studies using Van Veen grab to sample sediment for HTS assessments of eukaryotic diversity often randomly collect 1.5–2 g of sediment from across the top 1 cm surface layer of the grab to address the issue of micro-scale patchiness<sup>23,39–41,51–53</sup>. Unfortunately, we were not able to test in our experiment sample volumes of 1.5–2 g, as the two DNA extraction kits here used that contained the same reagents, were optimized for sample volumes not smaller than 4–5 g (PowerMax<sup>®</sup> Soil DNA Isolation Kit) or larger than 0.2 g (PowerSoil<sup>®</sup> DNA Isolation Kit). To test a sample volume of 1.5–2 g efficiently we would need to use a separate extraction kit with different protocols and reagents, making the comparison with remaining sample volumes difficult. Testing these often used sample volumes in future studies would provide valuable information.

**Community composition.** We found a high diversity of eukaryotes in brackish sediments of the Baltic Sea. In terms of relative OTU abundance (Fig. 6 and Supplement Fig. S6) the Copepoda was the most represented group, despite nematode being generally the most abundant taxon in metazoan datasets for brackish and marine sediments<sup>7,12,54</sup>. It is possible that the higher prevalence of copepod OTUs here found is related to: (a) the degenerate nature of the eukaryotic primers sets used in our study<sup>10,55</sup>, (b) high biomass of copepod tissue that was potentially differentially amplified<sup>56</sup> and (c) the extraction strategy used in our study designed to capture total (both intra- and extracellular) DNA. As such, and taking into account the size of our amplicons (~400 bp), our sample likely reflects DNA from contemporary and recently inhabiting organisms, in addition to DNA from planktonic organisms that has settled to the benthos. Copepods species such as *Acartia bifilosa* and *Acartia tonsa* found in our dataset deploy sediment resting stages<sup>57</sup>, thereby suggesting marine snow as the likely source of copepod eDNA. Similarly, Guardiola *et al.*<sup>37</sup> recovered more Arthropoda diversity compared to nematodes in 18 S nSSU



HTS analyses from samples obtained from deep sea canyons where eDNA was extracted directly from sediment. Nevertheless, our study detected 194 unique nematode OTUs suggesting that we did manage to effectively sample metazoan diversity (Supplement Fig. S10).

**Concluding remarks.** Collectively our work indicates that small spatial variability, even within a sampling unit, is a factor that needs to be taken into account in studies investigating eukaryotic and metazoan diversity in aquatic sediments. To achieve a representative assessment of the beta-diversity for both these communities it is necessary to sample a large percentage of the sampling unit (78% or 22 g). We saw similar pattern for alpha-diversity metrics (Table 1, Fig. 3B,D,F) but the different estimates of richness for the 100% treatment stop being significant at around 60% of the sampling unit. As such, if one is targeting solely metazoans it is probably more efficient to first separate the metazoan community from the sediment particles using elutriate or density extraction techniques<sup>4,14</sup> before DNA extraction as suggested by Brannock & Halanych<sup>12</sup>. Furthermore, when larger sampling units are to be used (e.g. Van Veen grabs) to simultaneously sample multiple size fractions of sediment communities, combining multiple sediment subsamples for microeukaryotes with sediment sieving and isolation of larger taxa has been shown to be effective<sup>23,41</sup>.

Nevertheless, it is clear from our study that small sample volumes of 0.2 g (1% of the sampled area), often used in benthic surveys to address sediment benthic diversity are not large enough to account for spatial heterogeneity, even at small spatial scales, and measure reliably both alpha and beta-diversity in non-metazoan eukaryotic communities. In contrast to other studies with marine sediments we did find a strong effect of sample volumes on alpha-diversity metrics for non-metazoan eukaryotes indicating a need for the use of larger sample volumes. In addition, our data suggest that approximately 14 g of homogenized sediment in our study needs to be sampled to achieve an adequate beta-diversity measurement.

The larger sample volumes here discussed seem to capture adequately the complex heterogeneous matrix of microhabitats that soft sediments provide for both micro metazoans and eukaryotes. In addition, our results are based on the use of a sampling core of one size, albeit a common one used in studies of micrometazoans<sup>3–5</sup> with sediment from one location in a coastal muddy soft-sediment habitat. It would be valuable to test these recommendations with additional sample volumes commonly used in field studies, and different locations with sediment of different types collected with sampling equipment of larger sizes.

Our results are relevant for both field and experimental studies in marine systems, but the present investigation did not aim to replicate current sampling strategies used in ongoing benthic monitoring initiatives (e.g. in terms of number of replicates per site and overall number of sample sites). However, our findings indicate that DNA extraction volume can significantly impact biodiversity estimates derived from metabarcoding approaches, and this should be taken into account when designing future DNA-based environmental monitoring strategies.

Our study underlines the need for a robust standardization of methods to investigate the diversity of eukaryotic communities and allow for cross-study comparisons, particularly taking into account the foreseeable future where DNA techniques will become more important in efforts to assess the dynamics of biodiversity in aquatic sediments<sup>58</sup>.

## Material and Methods

**Sample collection.** Silt muddy sediment was collected with five handheld Perspex sediment cores taken approximately 40–50 cm from each other from a depth of 15 m at a temperature of 5 °C and salinity of 6 psu in the Stockholm archipelago (58°50'N, 17°32'E) in the northern Baltic Sea, Sweden. The handheld sampling units were 44 mm diameter with a surface area of 15 cm<sup>2</sup>, a size which has been found to be appropriate for sampling of microbial benthic metazoans such as meiofauna<sup>59</sup>. After collection, the sediment cores were transported to Stockholm University and placed in a thermoconstant room at 5 ± 1 °C overnight. Previous studies report that >95% of the diversity of Baltic Sea metazoan diversity is found within the first centimetre of sediment<sup>60</sup>. As such, on the following day, the top 2 centimetres of sediment from each replicate sediment core were sliced, homogenized, weighed and divided into five sample volumes: 0.2 g, 4 g, 6 g, 8 g, and the remaining amount of sediment that varied between 10–11 g, corresponding to 1, 14, 21, 28 and 36% of the top 2 centimetres of each core. In this way, all the sediment from the top 2 cm of each of the five replicates was analysed, with the experimental design facilitating the investigation of the effect of sample volume in the composite core. The sediment samples were then placed in 50 mL Falcon tubes and kept at –20 °C until DNA extraction, with the exception of the 0.2 g samples that were placed in 1.5 mL Eppendorf tubes. The sample volumes were chosen taking into account the practical limitations of existing DNA extraction commercial kits and the need to have a sufficiently large number of different sediment quantities to test the effect of sample volume while looking at the whole surface area of the sampling unit at an affordable cost.

**Sample preparation and sequencing.** For the 0.2 g samples, DNA was extracted with the PowerSoil<sup>®</sup> DNA Isolation Kit (MOBIO, Cat#12888), following the protocol instructions. The PowerMax<sup>®</sup> Soil DNA Isolation Kit (MOBIO, Cat#12988) was used for the remaining sample volumes (4 g, 6 g, 8 g and 10 g). The protocol, methodology and chemical reagents used in the different kits are identical, the only difference being the amount of sediment each kit can process<sup>12</sup>. After DNA extraction, samples were frozen at –20 °C in 100 µL (0.2 g samples) and 3 mL (all other sample volumes of C6 solution (10 mM Tris)). Following this, 100 µL of each DNA extract was purified using PowerClean<sup>®</sup> Pro DNA Clean-Up Kit (MOBIO, Cat# 12997–50) and stored in 100 µL of C5 (10 mM tris) solution at –20 °C. DNA extraction from different sample volumes yielded DNA extracts of different concentration ranging from 62.6 to 10.1 ng/µL. Before PCR amplification all DNA extracts were standardized to a concentration of 10 ng/µL.

The highly conservative metabarcoding primers TAREuk454FWD1 (5'-CCAGCA(G/C)C(C/T)GCGTAATTCC-3') and TAREukREV3 (5'-ACTTTCGTTCTTGAT(C/T)(A/G)A-3')<sup>61</sup> and Pfu DNA

polymerase (Promega, Southampton, UK) were used to PCR amplify the 18S rDNA gene region, yielding fragments between 365–410 bp not including adaptors or barcodes. Each sample volume from the five replicate sampling cores were amplified in triplicates. These triplicates were then pooled, dual-barcoded with Nextera XT index primers following Bista *et al.*<sup>62</sup> and visualized by gel electrophoresis. These tagged amplicons were later purified with the Agencourt AMPure XP PCR Purification kit (Beckman Coulter), quantified with Qubit (Invitrogen, USA) and pooled in equimolar quantities. The purified amplicons were then sequenced in both directions on an Illumina MiSeq platform at National Genomics Infrastructure (NGI-Stockholm, Sweden) as a single pool comprised of the 25 different samples with 25 unique index primer combinations (i.e., five sample volumes – 0.2 g, 4 g, 6 g, 8 g, and 10 g – from 5 different sampling core replicates).

**Bioinformatics.** Amplicon reads were demultiplexed by NGI and raw demultiplexed reads were used in the initial data processing and quality-filtering was carried out in QIIME 1.8<sup>63</sup>. Raw Illumina reads were first subjected to read pairing (merging Read 1 and Read 2) using the `join_paired_ends.py` in QIIME, followed by a demultiplexing step using the `multiple_split_libraries_fastq.py` script which split samples according to dual-index barcodes and also performed quality-filtering of raw reads. Remaining forward and reverse PCR primer sequences were subsequently removed from Illumina reads using Trimmomatic version 0.32<sup>64</sup>. Processed reads were concatenated into a single file, and then subjected to open-reference OTU picking with Uclust using the `pick_open_reference_otus.py` script in QIIME (with 10% subsampling, no prefiltering, and reverse strand match enabled). Singleton and doubleton OTUs were discarded from the resulting OTU tables (retaining only OTUs with >2 reads). Two independent OTU picking workflows were carried out, using pairwise cutoffs at 96% and 99% sequence identity (hereby referred to as 96% and 99% OTUs). Taxonomy was assigned to representative OTU sequences using the RDP Classifier<sup>65</sup> in QIIME (`assign_taxonomy.py`), using the SILVA 119 release as a reference database<sup>66</sup>. OTU representative sequences were aligned using PYNAST<sup>67</sup> in `align_seqs.py` and used to construct a Phylogenetic tree using FastTree<sup>68</sup> in `make_phylogeny.py`. All bioinformatics commands, OTU tables, and quality-filtered Illumina reads have been posted on FigShare (doi: 10.6084/m9.figshare.4993667).

**Community analyses.** The number of reads for the five sample volumes (0.2 g, 4 g, 6 g, 8 g and 10 g) of the same replicate were then added in all possible combinations to produce nine additional sample volumes (12 g, 14 g, 16 g, 18 g, 20 g, 22 g, 24 g, 28 g and the complete 28.2 g) for the five replicate cores, corresponding to 43, 50, 57, 64, 71, 78, 85, 99 and 100% of the sediment in the top 2 cm of each core (see Supplement Table S5). This data set was imported into R v 3.3.2 (R Core Team 2016) and analysed using the *phyloseq*<sup>69</sup> and *vegan*<sup>70</sup> packages. Raw OTU abundances were transformed to relative abundances, which has been suggested as a more appropriate and effective alternative to rarefying<sup>48</sup>. The effect of sample volume on proportion of metazoan taxa, number of unique OTUs, Margalef, and ACE index were tested with one-way analysis of variance (ANOVA). ACE is commonly used as an alpha-diversity measure in HTS investigations that takes into account rare taxa, while the choice of the Margalef index was based on its high sensitivity to sample volume and good capability of discriminating between samples with different alpha-diversity<sup>71</sup>. Homogeneity of variances between groups were tested using Bartlett's test to examine for all ANOVA analysis. Differences among treatments, i.e. sample volumes, in relation to the 100% treatment were tested with a Tukey HSD *posthoc* test for all variables and were considered significant at a  $p < 0.01$ .

Community composition was examined by first dividing OTUs into Metazoa, and non-metazoan Eukaryotes groups. The dissimilarity between faunal assemblages in the different sample volumes was analysed by non-metric multidimensional scaling (NMDS), using the Sørensen's dissimilarity coefficient, and Principal Coordinates Analysis (PCoA) using Bray distances. To test for effects of sample volume on beta-diversity we conducted a PERMANOVA multivariate analysis of variance with the *adonis* function of the *vegan* package<sup>70</sup>. The function PERMDISP was used to test for differences in beta-diversity among sample volumes in relation to the 100% treatment. Additionally, we used metrics to partition beta-diversity to quantify the relative importance of turnover and nestedness in the different sample volumes<sup>72,73</sup>. These authors propose that beta-diversity measured as Sørensen dissimilarity index ( $\beta_{sør}$ ) can be separated into dissimilarity as a consequence of turnover, i.e. species replacement between sites or samples ( $\beta_{sim}$ ), and dissimilarity attributable to nestedness, species loss from sample to sample ( $\beta_{nes}$ ). We used the R package *betapart*<sup>74</sup> for this analysis. Following this, we compared how both components of beta-diversity, turnover and nestedness changed with sample volume in relation to the 100% treatment. All statistical tests were performed on R v 3.3.2.

**Data Accessibility.** Data associated with this manuscript is deposited on FigShare (doi:10.6084/m9.figshare.4993667).

## References

- Creer, S. *et al.* Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. *Mol. Ecol.* **19**(Suppl 1), 4–20 (2010).
- Leray, M. & Knowlton, N. DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proc. Natl. Acad. Sci. USA* **112**, 2076–81 (2015).
- Näslund, J., Nascimento, F. J. A. & Gunnarsson, J. S. Meiofauna reduces bacterial mineralization of naphthalene in marine sediment. *ISME J* **4**, 1421–1430 (2010).
- Nascimento, F., Näslund, J. & Elmgren, R. Meiofauna enhances organic matter mineralization in soft sediment ecosystems. *Limnol Ocean.* **57**, 338–346 (2012).
- Bonaglia, S., Nascimento, F. J. A., Bartoli, M., Klawonn, I. & Bruchert, V. Meiofauna increases bacterial denitrification in marine sediments. *Nat. Commun.* **5**, 5133 (2014).
- Danovaro, R., Snelgrove, P. V. R. & Tyler, P. Challenging the paradigms of deep-sea ecology. *Trends Ecol. Evol.* **29**, 465–475 (2014).
- Fonseca, V. G. *et al.* Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nat. Commun.* **1**, 98 (2010).

8. Bik, H. M. *et al.* Sequencing our way towards understanding global eukaryotic biodiversity. *Trends Ecol. Evol.* **27**, 233–43 (2012).
9. Leray, M. & Knowlton, N. Censusing marine eukaryotic diversity in the twenty-first century. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **371**, (2016).
10. Creer, S. *et al.* The ecologist's field guide to sequence-based identification of biodiversity. *Methods Ecol. Evol.* **7**, 1008–1018 (2016).
11. Taberlet, P., Coissac, E., Hajibabei, M. & Rieseberg, L. H. Environmental DNA. *Mol. Ecol.* **21**, 1789–1793 (2012).
12. Brannock, P. M. & Halanych, K. M. Meiofaunal community analysis by high-throughput sequencing: Comparison of extraction, quality filtering, and clustering methods. *Mar. Genomics* **23**, 67–75 (2015).
13. Sogin, M. L. *et al.* Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc. Natl. Acad. Sci. USA* **103**, 12115–20 (2006).
14. Giere, O. *Meiobenthology: The microscopic motile fauna of aquatic sediments.* (Springer-Verlag, 2009).
15. Hewitt, J. E., Thrush, S. F., Halliday, J. & Duffy, C. The importance of small-scale habitat structure for maintaining beta diversity. *Ecology* **86**, 1619–1626 (2005).
16. Yousef, N. H., Couger, M. B. & Elshahed, M. S. Fine-Scale Bacterial Beta Diversity within a Complex Ecosystem (Zodletone Spring, OK, USA): The Role of the Rare Biosphere. *PLoS One* **5**, e12414 (2010).
17. Kang, S. & Mills, A. L. The effect of sample size in studies of soil microbial community structure. *J. Microbiol. Methods* **66**, 242–250 (2006).
18. Takahara, T., Minamoto, T., Yamanaka, H., Doi, H. & Kawabata, Z. Estimation of Fish Biomass Using Environmental DNA. *PLoS One* **7**, e35868 (2012).
19. Torti, A., Lever, M. A. & Jørgensen, B. B. Origin, dynamics, and implications of extracellular DNA pools in marine sediments. *Mar. Genomics* **24**, 185–196 (2015).
20. Lorenz, M. G. & Wackernagel, W. Bacterial gene transfer by natural genetic transformation in the environment. **58**, 563–602 (1994).
21. Goldberg, C. S. *et al.* Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods Ecol. Evol.* **7**, 1299–1307 (2016).
22. Lanzén, A., Lekang, K., Jonassen, I., Thompson, E. M. & Troedsson, C. DNA extraction replicates improve diversity and compositional dissimilarity in metabarcoding of eukaryotes in marine sediments. *PLoS One* **12**, e0179443 (2017).
23. Stoeck, T., Kochems, R., Forster, D., Lejzerowicz, F. & Pawlowski, J. Metabarcoding of benthic ciliate communities shows high potential for environmental monitoring in salmon aquaculture. *Ecol. Indic.* **85**, 153–164 (2018).
24. Zhao, F. & Xu, K. Molecular diversity and distribution pattern of ciliates in sediments from deep-sea hydrothermal vents in the Okinawa Trough and adjacent sea areas. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **116**, 22–32 (2016).
25. Pasulka, A. L. *et al.* Microbial eukaryotic distributions and diversity patterns in a deep-sea methane seep ecosystem. *Environ. Microbiol.* **18**, 3022–3043 (2016).
26. Volant, A. *et al.* Spatial Distribution of Eukaryotic Communities Using High-Throughput Sequencing Along a Pollution Gradient in the Arsenic-Rich Creek Sediments of Carnoulès Mine, France. *Microb. Ecol.* **72**, 608–620 (2016).
27. Team, R. C. R: A language and environment for statistical computing (2016).
28. Bik, H. M., Halanych, K. M., Sharma, J. & Thomas, W. K. Dramatic Shifts in Benthic Microbial Eukaryote Communities following the Deepwater Horizon Oil Spill. *PLoS One* **7**, e38550 (2012).
29. Lallias, D. *et al.* Environmental metabarcoding reveals heterogeneous drivers of microbial eukaryote diversity in contrasting estuarine ecosystems. *ISME J.* <https://doi.org/10.1038/ismej.2014.213> (2014).
30. Aguilar, M. *et al.* Next-Generation Sequencing Assessment of Eukaryotic Diversity in Oil Sands Tailings Ponds Sediments and Surface Water. *J. Eukaryot. Microbiol.* **63**, 732–743 (2016).
31. Chariton, A. A. *et al.* A molecular-based approach for examining responses of eukaryotes in microcosms to contaminant-spiked estuarine sediments. *Environ. Toxicol. Chem.* **33**, 359–369 (2014).
32. Fonseca, V. G. *et al.* Revealing higher than expected meiofaunal diversity in Antarctic sediments: a metabarcoding approach. *Sci. Rep.* **7**, 6094 (2017).
33. Lecroq, B. *et al.* Ultra-deep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments. *Proc. Natl. Acad. Sci. USA* **108**, 13177–82 (2011).
34. Lanzén, A., Lekang, K., Jonassen, I., Thompson, E. M. & Troedsson, C. High-throughput metabarcoding of eukaryotic diversity for environmental monitoring of offshore oil-drilling activities. *Mol. Ecol.* **25**, 4392–4406 (2016).
35. Bhadury, P. & Austen, M. C. Barcoding marine nematodes: an improved set of nematode 18S rRNA primers to overcome eukaryotic co-interference. *Hydrobiologia* **641**, 245–251 (2010).
36. Smith, K. F., Kohli, G. S., Murray, S. A. & Rhodes, L. L. Assessment of the metabarcoding approach for community analysis of benthic-epiphytic dinoflagellates using mock communities. *New Zeal. J. Mar. Freshw. Res.* **51**, 555–576 (2017).
37. Guardiola, M. *et al.* Spatio-temporal monitoring of deep-sea communities using metabarcoding of sediment DNA and RNA. *PeerJ* **4**, e2807 (2016).
38. Pawlowski, J., Esling, P., Lejzerowicz, F., Cedhagen, T. & Wilding, T. A. Environmental monitoring through protist next-generation sequencing metabarcoding: assessing the impact of fish farming on benthic foraminifera communities. *Mol. Ecol. Resour.* **14**, 1129–1140 (2014).
39. Keeley, N., Wood, S. A. & Pochon, X. Development and preliminary validation of a multi-trophic metabarcoding biotic index for monitoring benthic organic enrichment. *Ecol. Indic.* **85**, 1044–1057 (2018).
40. Laroche, O. *et al.* A cross-taxa study using environmental DNA/RNA metabarcoding to measure biological impacts of offshore oil and gas drilling and production operations. *Mar. Pollut. Bull.* **127**, 97–107 (2018).
41. Lejzerowicz, F. *et al.* High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems. *Sci. Rep.* **5**, 13932 (2015).
42. Beck, J., Holloway, J. D. & Schwanghart, W. Undersampling and the measurement of beta diversity. *Methods Ecol. Evol.* **4**, 370–382 (2013).
43. Christie, P. & Beattie, J. A. M. Significance of sample size in measurement of soil microbial biomass by the chloroform fumigation-incubation method. *Soil Biol. Biochem.* **19**, 149–152 (1987).
44. Ellingsøe, P. & Johnsen, K. Influence of soil sample sizes on the assessment of bacterial community structure. *Soil Biol. Biochem.* **34**, 1701–1707 (2002).
45. Ranjard, L. & Richaume, A. Quantitative and qualitative microscale distribution of bacteria in soil. *Res. Microbiol.* **152**, 707–716 (2001).
46. Penton, C. R., Gupta, V. V. S. R., Yu, J. & Tiedje, J. M. Size Matters: Assessing Optimum Soil Sample Size for Fungal and Bacterial Community Structure Analyses Using High Throughput Sequencing of rRNA Gene Amplicons. *Front. Microbiol.* **7**, 824 (2016).
47. Eleftheriou, A. & McIntyre, A. *Methods for the Study of Marine Benthos. Methods for the Study of Marine Benthos: Third Edition.* <https://doi.org/10.1002/9780470995129> (2007).
48. McMurdie, P. J. & Holmes, S. W. Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Comput. Biol.* **10**, e1003531 (2014).
49. Nascimento, F. J. A., Karlson, A. M. L. & Elmgren, R. Settling blooms of filamentous cyanobacteria as food for meiofauna assemblages. *Limnol. Ocean.* **53**, 2636–2643 (2008).
50. Aylagas, E., Mendibil, I., Borja, Á. & Rodríguez-Ezpeleta, N. Marine Sediment Sample Pre-processing for Macroinvertebrates Metabarcoding: Mechanical Enrichment and Homogenization. *Front. Mar. Sci.* **3**, 203 (2016).

51. Chariton, A. A. *et al.* Metabarcoding of benthic eukaryote communities predicts the ecological condition of estuaries. *Environ. Pollut.* **203**, 165–174 (2015).
52. Pochon, X. *et al.* Accurate assessment of the impact of salmon farming on benthic sediment enrichment using foraminiferal metabarcoding. *Mar. Pollut. Bull.* **100**, 370–382 (2015).
53. Laroche, O. *et al.* First evaluation of foraminiferal metabarcoding for monitoring environmental impact from an offshore oil drilling site. *Mar. Environ. Res.* **120**, 225–235 (2016).
54. Creer, S. & Sinniger, F. Cosmopolitanism of microbial eukaryotes in the global deep seas. *Mol. Ecol.* **21**, 1033–1035 (2012).
55. Piñol, J., Mir, G., Gomez-Polo, P. & Agustí, N. Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Mol. Ecol. Resour.* **15**, 819–830 (2015).
56. Kaneryd, L. *et al.* Species-rich ecosystems are vulnerable to cascading extinctions in an increasingly variable world. *Ecol. Evol.* **2**, 858–874 (2012).
57. Katajisto, T., Viitasalo, M. & Koski, M. Seasonal occurrence and hatching of calanoid eggs in sediments of the northern Baltic Sea. *Mar. Ecol. Prog. Ser.* **163**, 133–143 (1998).
58. Leese, F. *et al.* DNAqua-Net: Developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe. *Res. Ideas Outcomes* **2**, e11321 (2016).
59. Montagna, P. A., Baguley, J. G., Hsiang, C.-Y. & Reuscher, M. G. Comparison of sampling methods for deep-sea infauna. *Limnol. Oceanogr. Methods* **15**, 166–183 (2017).
60. Ólafsson, E., Modig, H. & van de Bund, W. J. Species specific uptake of radio-labelled phyto-detritus by benthic meiofauna from the Baltic Sea. *Mar. Ecol. Prog. Ser.* **177**, 63–72 (1999).
61. Stoeck, T. *et al.* Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* **19**, 21–31 (2010).
62. Bista, I. *et al.* Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. *Nat. Commun.* **8**, 14087 (2017).
63. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
64. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–20 (2014).
65. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
66. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
67. Caporaso, J. G. *et al.* PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**, 266–267 (2010).
68. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
69. McMurdie, P. J. & Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* **8**, e61217 (2013).
70. Oksanen, A. J. *et al.* Vegan: Community Ecology Package. <https://cran.r-project.org>, <https://github.com/vegandevs/vegan> 291 (2016).
71. Magurran, A. E. In *Ecological Diversity and Its Measurement* 61–80 (Springer Netherlands). [https://doi.org/10.1007/978-94-015-7358-0\\_4](https://doi.org/10.1007/978-94-015-7358-0_4) (1988).
72. Baselga, A. Partitioning the turnover and nestedness components of beta diversity. *Glob. Ecol. Biogeogr.* **19**, 134–143 (2010).
73. Baselga, A. & Leprieux, F. Comparing methods to separate components of beta diversity. *Methods Ecol. Evol.* **6**, 1069–1079 (2015).
74. Baselga, A. & Orme, C. D. L. betapart: an R package for the study of beta diversity. *Methods Ecol. Evol.* **3**, 808–812 (2012).

## Acknowledgements

We would like to thank three anonymous reviewers for comments and suggestion that have helped to improve the manuscript, Caroline Raymond for assistance with sediment collection and the MEFGL staff for help with the laboratory work. F.N.'s participation in this project was supported by the Swedish Research Council (grant nr 623-2010-6616), the Swedish Research Council Formas (Mobility Grant 2013-1322) and the Lars Hierta Minne Foundation. Sequencing was performed at the National Genomics Institute, Sweden, with the support of the SciLifeLab National Project in Genomics.

## Author Contributions

F.N. and S.C. conceived the ideas and designed the experimental plan; F.N. and D.L. performed the laboratory analysis; H.B. led the bioinformatics work and F.N. and S.C. analysed and interpreted the data. While F.N. led the writing, all authors contributed the writing of the manuscript with critical input and gave final approval for publication.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-30179-1>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018