Advances in metabarcoding techniques bring us closer to reliable monitoring of the marine benthos

Steyaert, Margaux ; Priestley, Victoria ; Osborne, Owen; Herraiz, Alba ; Arnold, Richard; Savolainen, Vincent

Journal of Applied Ecology

DOI:
10.1111/1365-2664.13729

Published: 01/11/2020

Peer reviewed version

Title: Advances in metabarcoding techniques bring us closer to reliable monitoring of the marine benthos

Authors: Margaux Steyaert[1,2], Victoria Priestley[1], Owen Osborne[1,3], Alba Herraiz[1], Richard Arnold[4], Vincent Savolainen[1]*

[1] Department of Life Sciences, Imperial College London, Silwood Park Campus, SL5 7PY, United Kingdom.

[2] Current address: Department of Zoology, University of Oxford, Zoology Research and Administration Building, 11a Mansfield Road, Oxford, OX1 3SZ.

[3] Current address: School of Natural Sciences, Bangor University, Bangor, Gwynedd, LL57 2DG, United Kingdom.

[4] Thomson Environmental Consultants, Surrey Research Park, GU2 7AG, United Kingdom.

*Corresponding author: v.savolainen@imperial.ac.uk

Abstract

1. Reliable and accurate biodiversity census methods are essential for monitoring ecosystem health and assessing potential ecological impacts of future development projects. Although metabarcoding is increasingly used to study biodiversity across ecological research, morphology-based identification remains the preferred approach for marine ecological impact assessments. Comparing metabarcoding to morphology-based protocols currently used by ecological surveyors is essential to determine whether this DNA-based approach is suitable for the long-term monitoring of marine ecosystems.

2. We compared metabarcoding and morphology-based approaches for the analysis of invertebrates in low diversity intertidal marine sediment samples. We used a recently developed bioinformatics pipeline and two taxonomic assignment methods to resolve and assign amplicon sequence variants (ASVs) from Illumina amplicon data. We analysed the community composition recovered by both methods and tested the effects, on the levels of diversity detected by the metabarcoding method, of sieving samples prior to DNA extraction.

3. Metabarcoding of the mitochondrial marker cytochrome c oxidase I (COI) gene recovers the presence of more taxonomic groups than the morphological approach. We found that sieving samples results in lower alpha diversity detected and suggests a community composition that differs significantly from that suggested by un-sieved samples in our metabarcoding analysis. We found that whilst metabarcoding and morphological approaches detected similar numbers of species, they are unable to identify the same set of species across samples.

4. *Synthesis and Applications* We show that metabarcoding using the COI marker provides a more holistic, community-based, analysis of benthic invertebrate diversity than a traditional morphological approach. We also highlight current gaps in reference databases and bioinformatic pipelines for the identification of intertidal benthic invertebrates that need to be addressed before metabarcoding can replace traditional methods. Ultimately, with these limitations taken into consideration, resolving community-wide diversity patterns with metabarcoding could improve the management of non-protected marine habitats in the U.K.

Keywords: Biodiversity, Biomonitoring, Marine Benthos, Metabarcoding, Ecological Surveying.

Introduction

Understanding and quantifying the diversity of organisms is fundamental in the assessment of ecosystem health. Detecting significant shifts in species composition can lead to important changes in environmental policy, conservation efforts or the management of wild resources. As ecosystems are increasingly under pressure from climate and land-use change, it is vital that we understand which species are present or absent in habitats that interact with human civilization (Bardgett & van der Putten, 2014; Cardinale et al., 2012). In turn, by understanding trends in species composition, we can better quantify the value of ecosystems and the services they provide (Hautier et al., 2015). Reliable and fast methods for surveying species diversity are therefore highly sought after within both academia and the public sector (Keck et al., 2017; Baird et al., 2012).

Marine invertebrates have long been used to categorise and assess the health of marine ecosystems as shifts in their composition often reflect on wider patterns of human impact or natural disturbances (Borja, 2019; Chain et al., 2016). These organisms are considered important ecosystem bioindicators and have been utilised to screen the level of pollution and other anthropogenic impacts on marine habitats (Pérez et al., 2019; Poikane et al., 2016; Chiarelli & Roccheri, 2014). Macrobenthic invertebrates form a vital component of current biomonitoring programs, such as the European Union Water and Marine Strategy Framework Initiative (Hoey et al., 2019). Littoral and estuarine zones are key habitats often harbouring economically and ecologically important species but are increasingly impacted by the expansion of human development and pollution. In the United Kingdom, impact assessments following Water Framework Directive guidelines are required to be submitted when planning infrastructure development projects along the coastline in order to assess the level of impact such activities may cause (Environmental Agency, 2016). Such surveys routinely include an evaluation of macrobenthic invertebrate diversity, along with analyses of sediment particle size and isotopes. To our knowledge, all UK-based ecological consultancy companies currently offering marine consultancy services, including Environmental Impact Assessment (EIA) or Habitat Regulations Appraisals (HRA) surveys, only use traditional census methods in their identification of marine benthic invertebrates. These methods rely on examining morphological traits using light microscopy to taxonomically identify species and have been widely used to study macroinvertebrate diversity. A significant advantage to a morphological

approach is its ability to distinguish organisms that are present in a sample from biological remnants of transient species as well as enabling a direct count of individuals.

DNA metabarcoding enables the bulk identification of multiple species within an ecological sample by simultaneously amplifying individual 'DNA barcodes' (that is, DNA fragments that can be used for species identification), which are then sequenced and identified using HTS. This genetic method can allow for the identification of organisms that are too small or too degraded for light microscopy protocols, as well as cryptic taxa or species that exhibit phenotypic plasticity (Elbrecht & Leese, 2015). Finally, metabarcoding allows species diversity to be observed over a large spatial and temporal window, since genetic material from both present and transient organisms can be detected (Leray & Knowlton, 2015; Thomsen & Willerslev, 2015). However, there are still several limiting factors that prevent it from completely replacing traditional methods (Kelly et al., 2017; Lejzerowicz et al., 2015). These include a lack of available reference sequences in genetic databases, primer bias for amplification, copy number variation in target loci as well as unstandardized sample processing and sequence data analysis steps (Elbrecht et al., 2017; Drummond et al., 2015).

Metabarcoding is currently gaining considerable popularity, with many studies having successfully recovered the presence and diversity of marine species using this novel tool (Pearman et al., 2018; Yamamoto et al., 2017; Chain et al., 2016; Chariton et al., 2015). Several studies have directly compared traditional and metabarcoding approaches to surveying local marine benthic diversity (Aylagas et al., 2018; Cahill et al., 2018; Lobo et al., 2017; Lejzerowicz et al., 2015). A metabarcoding approach to sampling benthic macroinvertebrates has been shown to outperform traditional methods in the level of diversity recovered within an ecological sample (Lobo et al., 2017). However, datasets resulting from both methods are often difficult to compare directly as individuals are identified to different taxonomic levels (Cahill et al., 2018; Aylagas et al., 2016). Whilst there is a growing consensus that the future of biomonitoring now lies with high-throughput sequencing (HTS) methods such as metabarcoding and the targeting of environmental DNA (Aylagas et al., 2018; Pawlowski et al., 2018; Baird & Hajibabaei, 2012), further research must be undertaken prior to integrating these approaches into public sector biomonitoring. This includes comparing metabarcoding and morphology-based census methods currently used by companies offering ecological surveying services to evaluate how these protocols may differ from or complement one another.

In this study, we present a comparison between metabarcoding and morphological approaches for the assessment of species diversity in intertidal marine benthos samples. We follow a morphology-based protocol routinely used to survey marine macrobenthic diversity by a leading UK ecological consultant, Thomson Environmental Consultants. In order to directly compare each method's ability to detect and identify species, we perform both analyses on sets of environmental cores sampled from the same locations in an estuarine ecosystem. We hypothesise that our metabarcoding approach will 1) detect the presence of all macroinvertebrate species identified in the morphological approach and 2) will recover a larger range and diversity of organisms, including specimens only detectable via environmental DNA traces. We evaluate the effect of sieving versus not sieving samples prior to DNA extractions in order to assess the amount of organismal diversity represented by size fractions smaller than 0.5mm (a commonly used minimum size for morphology-based identification). Overall, this study benchmarks biomonitoring methods and provides further insight into the potential suitability of DNA based identification methods for the surveying of marine benthos communities.

Materials and methods

Sample collection

A total of 20 1-litre benthic samples were collected at 10 sites along the intertidal region of the Harwich International Port estuary (Norfolk, UK), in April 2017 (Figure 1, Table S1). This site is regularly surveyed by Thomson Environmental Consultants as part of an Environmental Impact Assessment project they carry out for the Harwich Haven Authority. All benthic cores were collected at low tide. For each sampling site, two cores (one for metabarcoding analysis and one for morphological identification) were extracted within 10cm of each other by inserting an extraction tube (surface area of 0.01 m^2) to a depth of 10cm. Cores collected for the metabarcoding analysis were placed in individual sterile Whirl-Pak® (Nasco, USA) bags and kept on ice during transport. Sterile gloves were always worn and replaced between each collection so as to limit cross contamination. Cores were stored at -80°C approximately 6 hours after field collection.

Sample processing, homogenisation and DNA extraction

Overall, all sample processing, DNA extractions, sequence amplification, library prep and sequencing stages were undertaken at Imperial College (see schematic overview in Figure S1). Prior to sample homogenisation and DNA extractions, cores were thawed at 4°C for 24 hours. The extraction apparatus was washed with nuclease-free water and detergents. Core samples number 1, 3, 5, 7 and 9 were individually sieved using a 0.5mm sterile mesh sieve. Organisms and biological matter were then separated from sediment using a decantation step whereby approximately 200g of benthos, along with 500ml of purified nuclease-free water, were first added to a 1L graduated cylinder, covered with Parafilm, and then were vigorously shaken before being decanted through the sieve. Empty shells were checked for sessile organisms and discarded prior to homogenisation. Remaining organic matter and organisms were collected and crushed using a sterile pestle and mortar. Core samples number 2, 4, 6, 8 and 10 were homogenised using a bulk blending approach. Cores were individually mixed in a sterile 1.5L glass blender (Klarstein, 700W) on the highest setting for 10 minutes. DNA was then extracted from two individual 8.5g technical replicate sub samples from each mixed or crushed core using the Mo Bio PowerMax® Soil DNA Isolation Kit (Qiagen), following the manufacturer's instructions. We extracted DNA from two technical replicates in order to conduct parallel polymerase chain reaction (PCR) and sequencing runs of each sample. Results from these technical replicates are then merged in the bioinformatics pipeline. Glassware were autoclaved and worktops bleached between each extraction to avoid cross contamination. The extracted DNA samples were then purified and concentrated using an ethanol precipitation protocol (Supplementary Text 2).

Morphological identification protocol

Thomson Environmental Consultants processed and analysed 20 cores following the National Marine Biological Analytical Quality Control Scheme (NMBAQC) Processing Requirements protocol for the identification of invertebrate species using light microscopy (Worsfold and Hall, 2010). Cores were filtered using a 0.5 mm meshed sieve. All organisms retained by the sieve were counted and identified to species level where possible by taxonomic experts.

Library preparation & sequencing

A 313 base pairs (bp) fragment of the COI gene was targeted using two universal primers with attached overhang Illumina adapters (*mlCOIintF* and *jgHCO2198*; Geller et al., 2013; Leray et al., 2013; Table S3). The amplicon region targeted by this degenerate primer pair

has been shown to be one of the most effective for metazoan metabarcoding, especially for the identification of marine macroinvertebrates (Ransome et al., 2017; Aylagas et al., 2016; Leray et al., 2013; Leray et al., 2015). Library preparation was carried out following recommendations made in Illumina's 16S Metagenomic Sequencing Library Preparation protocol (Illumina; Supplementary Text 1). This library was then sequenced on an Illumina Miseq platform using a MiSeq reagent kit v3 (2x300 cycle).

Sequence analysis

The open-source software package DADA2 (version 1.12) was used to quality check, filter, trim and remove chimeras from the raw demultiplexed reads following the online DADA2 Pipeline Tutorial 1.12 (https://benjjneb.github.io/dada2/tutorial.html) in R Studio 1.2.5019 (Callahan et al., 2016). DADA2 infers exact amplicon sequence variants (ASVs) from large amplicon datasets by creating and using a parametric error matrix. This enables biological sequences to be inferred prior to steps in the metabarcoding pipeline that can introduce errors from PCR and sequencing. Using ASV methods to analyse metabarcoding datasets have been shown to provide higher resolution of community composition than traditional Operational Taxonomic Unit (OTU) methods which are based on clustering sequencing reads based on a pre-determined dissimilarity threshold (Callahan et al., 2017; Needham et al., 2017). Taxonomy was then assigned to the resulting list of ASVs from DADA2 using the insect R package (version 1.3.0.9000) following the online tutorial https://cran.r-project.org/web/packages/insect/vignettes/insect-vignette.html) and a reference dataset made using the MIDORI-UNIQUE database specific to the *mlCOIintF_F/jghCO2198* primer amplicon region (Wilkinson et al., 2018). This latter package assigns taxon identification using classification trees. An alternative taxonomy assignment on the ASV list was also carried out using the online BLASTn tool (standard nucleotide BLAST) and the online nucleotide collection (Altschul et al., 1990).

The R package LULU 0.1.0, along with the command line package VSEARCH 2.14.2, were used to curate the ASV list from the DADA2 pipeline (Froezlev et al., 2017; Rognes et al., 2016). The online LULU R package tutorial was followed, along with recommended default settings (https://github.com/tobiasgf/lulu). LULU evaluates the co-occurrence of ASVs amongst samples and removes potential erroneous variants, resulting in more realistic diversity estimates and metrics (Froeslev et al., 2017). A step by step breakdown of the DADA2, insect and LULU pipelines is available as supplementary information (R

Harwich_metabarcoding_DADA2_LULU_script). The 'phyloseq' R package 1.30.0 was used to visualise the taxonomic composition and estimate the alpha diversity of samples, following the online phyloseq tutorial (https://vaulot.github.io/tutorials/Phyloseq_tutorial.html) (McMurdie & Holmes, 2013). The packages 'vegan' (version 2.5.6) and 'DESeq2' (version 1.26.0) were then used to analyse beta diversity across sieved and un-sieved samples. These packages were used to run a beta-dispersion 'betadisper' to test for homogeneity of dispersion amongst sieved and un-sieved samples, a permutational ANOVA 'adonis' (with 999 random permutations) to test for significant differences in community composition between sieved and un-sieved samples, and to plot a Principle Coordinates Analysis (PCoA) to visualise community similarity between the two processing steps. A step by step breakdown of the 'phyloseq', 'vegan' and 'DESeq2' analyses is available as supplementary information (R Harwich_metabarcoding_data_analysis_script).

Results

Sequencing results and the effect of sieving on detecting species diversity

Sequencing resulted in a total of 22,146,908 raw reads (Table 1). Quality filtering, merging and chimera removal steps resulted in a total of 621,170 merged reads across all 20 technical replicates (Table 1). Overall, 87% of reads were removed during a strict DADA2 filtering step so as to avoid spurious results further on in the pipeline (Table 1). A total of 1,662 ASVs were identified across samples, including 1,405 belonging to marine taxonomic groups. Of these, 307 ASVs were identified across 13 metazoan marine phyla.

More ASVs were detected in un-sieved samples (1,509) than in sieved samples (734) (Figure 2). Sequence alpha rarefaction curves level off across all sieved and un-sieved samples, indicating that the majority of the estuarine diversity has been sampled by our metabarcoding approach (Figure 2). A permutational ANOVA test determined that the community structure in sieved and un-sieved samples were significantly different ($F = 1.24$, $p < 0.05$). Annelida was found to be the dominant phylum in both sieved and un-sieved samples (representing 64.9% and 64.6% of reads, respectively) (Figure 3). Molluscs made up a larger portion of un-sieved sample reads than sieved sample reads (27.2% and 19.4%, respectively). Annelid

species *Amphichaeta sannio, Paranais litoralis, Phyllodoce groenlandica*, copepod *Delavalia palustris* and colonial hydroid *Clava multicornis* were only found in un-sieved samples. The common cockle *Cerastoderma edule* and flatworm *Zonorhynchus seminascatus* species were only found in sieved samples. Two chordate taxa, *Homo sapiens* (humans) and *Astyanax* (blind cave fish), were detected across several samples. The presence of human DNA is either due to contamination during the sampling, processing or extraction steps or potentially due to the presence of sewage in the sampling location at Harwich International port. We believe the presence of blind cave fish DNA is due to lab contamination as another research project focusing on this species was taking place within the same laboratory during the time of this study's metabarcoding analysis.

Morphological analysis

A total of 2,144 specimens were identified across the 10 core samples. Sample 9 had the highest number of organisms (583 individuals) and sample 8 had the least (73 individuals) (Table S2). Specimens representing six different phyla, 14 orders and 24 families of macroinvertebrates were identified. A total of 25 species from six different phyla were identified (Table 2). On average, 78% of individual specimens (1,414 individuals) were identified down to species level across samples, with sample 2 having the highest identification rate and samples 8 having the lowest (97% and 44% respectively; Table S2). Platyhelminthes, nemerteans and nematodes were only identified to phylum level. Unidentified animal eggs were also detected in one sample. Annelid worms (*Tharyx*, *Tubificoides* and *Streblospio*) dominated total specimen counts and were present in all samples (Table S2).

Comparing morphological and metabarcoding datasets across samples

All seven phyla detected in the morphological analysis were found in the metabarcoding analysis (Table 2). Overall, a total of 24 marine species were identified across metabarcoding samples and 25 species were identified by the morphological approach (Table 2, although some more taxa were identified at higher taxonomic level). However, only 11 species were identified by both methods (Table 3). Several taxa that were identified by the morphological approach and not identified by the metabarcoding approach were found to have representative sequences in the MIDORI-UNIQUE reference dataset used in this study (Table 3). However, *Exogone naidina, Sphaerosyllis tetralix, Eusarsiella zostericola, Abra tenuis, Tharyx robustus* and *Tharyx killariensis* had no representative sequences in the MIDORI-

UNIQUE reference dataset. The BLASTn search using the DADA2 output file identified eight species that had been detected by the morphological approach, but not by the metabarcoding approach described above. Both morphology and DNA identified Annelida as the most common phylum (83% of specimens identified by the morphological approach and 65% of total metazoan reads in metabarcoding; Figure 5.). Morphological identified specimens from 15 families whilst the metabarcoding insect/RDP classifier and BLASTn taxonomy assignment recovered 10 and 9 families, respectively.

Discussion

Sieving is a method commonly utilized in metabarcoding studies surveying marine macrobenthic invertebrates to partition bulk samples, and smaller size fractions have often been found to be the most diverse (Pearman et al., 2018; Wangensteen et al., 2018; Ransome et al., 2017). We find that sieving samples in the metabarcoding analysis results in a reduction in the number of species identified (Table 2) as well as a reduction in amplicon sequence variant (ASV) richness and alpha diversity estimates (Figure 2), indicating that a large portion of reads originate either from whole organisms that are smaller than 0.5 mm or from environmental DNA. We find that whilst ASV richness and diversity estimates are higher in un-sieved samples, sieved samples had the highest number of reads (62,995 reads in sieved samples versus 44,221 reads in un-sieved samples). This is to be expected as the removal of sediment and fine inorganic matter would have concentrated the amount of biological tissue used for DNA extraction.

Our study shows that whilst all samples harbour low meiofauna diversity in general, sieving prior to DNA extraction also had a significant effect on the community composition recovered in metabarcoding samples. This is also to be expected as benthic meiofauna, which range between 40µm up to 500µm and form an important part of intertidal diversity (Coull & Chandler, 2001), would have been washed out in sieved samples.

Arthropods form a large and important component of marine zooplankton and benthos, often acting as key intermediates in food webs (Pearman & Irigoien, 2015). In the identification of marine arthropods, the morphological analysis only identified the presence of *Eusarsiella*

zostericola, a non-native myodocopid ostracod in samples 3, 7, 8 and 10. In comparison, metabarcoding recovered the presence of myodocopid ostracods, as well as calanoid copepods of the *Acartia* genus, across all samples. Furthermore, there were considerably more arthropod sequence reads across un-sieved samples than in sieved samples (1,578 and 483 reads, respectively), including reads identified as the copepod species *Delavalia palustris* in the unsieved sample 4. These organisms can range in size smaller or larger than 0.5mm, meaning some will have been washed away in sieving steps in both the morphological approach and in some metabarcoding samples.

The ability of metabarcoding to detect minute organisms is advantageous as it allows us to better understand the true diversity of intertidal marine benthos, unlike standard morphology-based approach, which is limited to surveying organisms larger than 500 µm. The presence and diversity of meiofauna communities have been shown to reflect patterns of environmental degradation and levels of pollution (Morad et al., 2017). Recovering the presence of both meio- and macro-fauna is therefore important when assessing the health of degraded areas such as the Harwich International Port. However, whilst metabarcoding allows for a more holistic community-based approach, we recommend that careful consideration be taken when deciding to implement sieving in metabarcoding protocols.

We find that overall the metabarcoding analysis recovered almost double the number of animal phyla than the morphological method (13 metazoan phyla in the metabarcoding analysis vs 7 metazoan phyla in the morphological approach). Metabarcoding was able to recover the presence of several marine species in phyla not targeted by the morphological approach, including the hydrozoan *Clava multicornis* and the kinorhynch *Pycnophyes kielensis*. Whilst metabarcoding recovered the presence of more taxonomic groups than the morphological approach, it appears the overall diversity of the Harwich International Port estuary is very low and has been effectively sampled as rarefaction curves level off in all samples.

Annelids form a major part of estuarine benthic ecosystems and are often the most abundant phylum of macroinvertebrates detected by COI metabarcoding studies (Haenel et al., 2017; Aylagas et al., 2016a). Furthermore, annelids have been shown to dominate estuarine mud-flat environments and are often used as indicator taxa for characterizing intertidal estuarine environments (Conde et al., 2013). Both methods detected the dominance of annelid worms

across all samples (Figure 5). The species found to have the highest number of reads across all metabarcoding samples is the carnivorous polychaete *Nephtys hombergii*. In fact, the two sequence variants with the highest abundance of reads across samples were both identified as *Nephtys hombergii*, indicating the presence of potential intraspecific genetic diversity of the gene region targeted by the *mlCOI_intF/jghCO2198* primer pair. Both morphological and metabarcoding approaches recover the presence of this species in samples 1 to 8, and not in samples 9 and 10. Similarly, both methods detected the presence of polychaete worms of the family Cirratulidae in the third transect (samples 9 and 10). Whilst both methods were able to recover matching ecological distributions of these two taxa, not all species identified in both analyses were detected in the same samples. For example, the metabarcoding analysis recovered the presence of the common polychaete *Hediste diversicolor* and saltwater clams *Macoma balthica* and *Nucula nitidosa* in several samples, more than the morphological approach. In contrast, the morphological analysis recovered polychaete worms of *Capitella* and *Streblospio* genera across more samples than the metabarcoding approach.

With metabarcoding, we recovered the presence of important UK indicator species such as *Hediste diversicolor* and *Scrobicularia plana*. However, as abundance is measured here by the number of sequence variants in metabarcoding, it is not possible to know whether a species is found in high abundance due to a large number of individual organisms detected or as a result of DNA extracted from a large number of cells. This highlights a current pitfall of metabarcoding methods, which cannot provide yet accurate estimates of abundance, which is needed in some common benthic indices (Borja, 2019; Conde et al., 2013).

Recent studies have suggested that analysing high-throughput amplicon sequencing data using amplicon sequence variants (ASVs), which involves resolving the sequenced region down to the level of individual nucleotides by estimating and applying modelled error rates, provides a more accurate representation of diversity than using traditional sequencing OTU clusters (Glassman & Martiny, 2018; Callahan et al., 2017). Whilst our study finds that using metabarcoding sequences results in a species count comparable with the morphological method (Table 2), not all the species found in the morphological approach are identified by the metabarcoding approach and vice versa (Table 3). Of the 36 taxonomic groups (species, genera or family group) identified in the morphological analysis, only 11 of these were detected by our metabarcoding analysis using the RDP classifier and the MIDORI-UNIQUE reference dataset. Sixteen of the taxa identified in the morphological analysis and not in the

metabarcoding analysis had representative sequences in the MIDORI-UNIQUE reference dataset. Our alternative BLASTn search, using the ASV sequences from the DADA2 step and the online nucleotide collection, was able to recover the presence of half of these missing taxa. Several species and genera, which are detected in the morphological analysis and have representative sequences in both the MIDORI-UNIQUE and online nucleotide database, remained un-identified by both the *insect* and BLASTn taxonomy assignment methods (Table 3). It is possible that these taxonomic groups were mis-identified in the morphological approach. It is also possible that were not identified as a result of the primer pair used in this study's metabarcoding approach. Only a limited set of representative annelid specimens were used to create the primer set used in this study and these originated from specimens collected for the Moorea Biocode project, an initiative which is based in French Polynesia and focuses on assembling specimens from tropical ecosystems (Leray et al., 2013). Previous metabarcoding research has described the difficulty of deriving species level taxonomic assignment for marine benthic fauna due to the paucity of reference barcode sequences in public databases along with the presence of mis-identified and erroneous sequences (Leray et al., 2015). In this study we show that whilst representative sequences are available for the majority of fauna found in our samples, potential primer bias likely played a part in the failure to recover the same set of species as the morphological approach. The use of multiple "barcode" genes, and the use of a more degenerate set of primers (for example the recently developed Leray-XT primer pair), are ways of reducing marker bias and allowing for improved representation of the species composition within an ecological sample (Wangensteen et al., 2018; Alberdi et al., 2017; Drummond et al., 2015).

With the advent of cheaper and faster HTS methods, metabarcoding has become economically viable and therefore attractive for businesses and governments to use as part of their ecological assessment protocols. Metabarcoding has already been used to detect shifts in macroinvertebrate composition around oil-drilling platforms and in response to land use change (Laroche et al., 2017; Beng et al., 2016; Lanzén et al., 2016). There is a now a growing consensus that the future of marine benthic biomonitoring lies with HTS methods, such as metabarcoding and the targeting of environmental DNA (Carvalho et al., 2019; Aylagas et al., 2018; Baird & Hajibabaei, 2012). Our study presents a comparison of a metabarcoding approach to a morphological protocol regularly used by a leading environmental consultancy firm. We demonstrate that metabarcoding allows for a more holistic, cross-community, approach that recovers the presence of meio- and macro-faunal

383 taxa across many more phyla groups than a morphological approach. Our findings show that
384 the use of different taxonomy-assignment methods and reference databases can lead to
385 inconsistent species-level identification in the metabarcoding analysis. Whilst bioinformatic
386 pipelines and analysis tools for HTS are constantly evolving and improving, there is still a
387 need for exploratory studies of understudied taxa such as marine benthic meio- and macro-
388 fauna. A way to tackle the current paucity of reference databases would be to encourage
389 environmental consultancy firms and the research community to archive and barcode
390 specimens collected during traditional morphometric surveys, so that localised curated
391 reference datasets may be built over time to facilitate future metabarcoding efforts.

Data Accessibility

Raw FASTQ files are available online (https://datadryad.org/stash/share/XfL_GJyDgvFKW1l3V8ihvWHXQXkxLyLCLJWh1-b2sQQ).

Authors' contributions

MS, VP and VS designed the project. VS and RA supervised the research. MS collected and analysed data. AH helped with the laboratory work. OGO helped with bioinformatics pipelines. MS wrote the initial manuscript with subsequent contributions from all authors.

References

Alberdi, A., Aizpurua, O., Thomas, M., Gilbert, P., Bohmann, K. (2017). Scrutizing key steps for reliable metabarcoding of environmental samples. Methods in Ecology and Evolution. 9, 134-147. https://doi.org/10.1111/2041-210X.12849

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Avó, A. P., Daniell, T. J., Neilson, R., Oliveira, S., Branco, J. & Adão, H. (2017). DNA barcoding and morphological identification of benthic nematodes assemblages of estuarine intertidal sediments: Advances in molecular tools for biodiversity assessment. Frontiers in Marine Science. 4, 66. https://doi.org/10.3389/fmars.2017.00066

Aylagas, E., Borja, A., Irigoien, X. & Rodriguez-Ezpeleta, N. (2016a). Benchmarking DNA metabarcoding for biodiversity-based monitoring and assessment. Frontiers in Marine Science. 3, 96. https://doi.org/10.3389/fmars.2016.00096

Aylagas E., Borja, A., Muxika, I., Rodriguez-Ezpeleta, N. (2018). Adapting metabarcoding-based benthic biomonitoring into routine ecological status assessment networks. Ecological Indicators. 95 (1), 194-202. https://doi.org/10.1016/j.ecolind.2018.07.044

Bardgett, R. D. & van Der Putten, W., H. (2014). Below ground biodiversity and ecosystem functioning. Nature. 515, 505-511. https:// doi:10.1038/nature13855

Beng, C. K., Tomlinson, K. W., Xian, H. S., Yann Surget-Groba, Hughes, A. C., Corlett, R. T. & Ferry Slik, J. W. (2016). The utility of DNA metabarcoding for studying the response of arthropod diversity and composition to land- use change in the tropics. Scientific Reports. 6, 24965. https://doi: 10.1038/srep24965

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

Borja, A. (2019) Benthic communities and their use in marine health monitoring programs, including new genomic tools. Conference: International Conference on Benthos (ICB 2019).

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., Holmes, S. P. (2016). DADA2: High resolution sample inference from Illumina amplicon data. Nature Methods. 13, 581-583. https://doi.org/10.1038/nmeth.3869

Callahan, B. J., McMurdie, P. J., Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. The ISME Journal. 11, 2639-2643. https://doi.org/10/1038/ismej.2017.119

Caporaso, J., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F., Costello, E., … Knight, R. (2010). QIIME allows analysis of high- throughput community sequencing data. Nature Methods. 7 (5), 335-6. https://doi.org/10.1038/nmeth.f.303

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. https://doi.org/10.1038/nature11148

Carvalho, S., Aylagas, E., Villalobos, R., Kattan, Y., Berumen, M., Pearman, J. K. (2019). Beyond the visual: using metabarcoding to characterize the hidden reef cryptobiome. Proceedings of the Royal Society B: Biological Sciences. 286 (1896), 20182697. https://doi.org/10.1098/rspb.2018.2697

Chain, F. J. J., Brown, E. A., Macisaac, H. J. & Cristescu, M. E. (2016). Metabarcoding reveals strong spatial structure and temporal turnover of zooplankton communities among marine and freshwater ports. Diversity and Distributions. 22 (5), 493-504. https://doi.org/10.1111/ddi.12427

Chariton, A. A., Stephenson, S., Morgan, M. J., Steven, A. D. L., Colloff, M. J., Court, L. N. & Hardy, C. M. (2015). Metabarcoding of benthic eukaryote communities predicts the ecological condition of estuaries. Environmental Pollution. 203, 165. https://doi.org/10.1016/j.envpol.2015.03.047

Chiarelli, R., Roccheri, M. C. (2014). Marine Invertebrates as bioindicators of heavy metal pollution. Open Journal of Metal. 4, 93-106. https://dx.doi.org/10.4236/ojmetal.2014.44011

Conde, A., Novais, J & Dominguez, J. (2013). Characterization of an estuarine environment by means of an index based on intertidal macrofauna. Marine Pollution Bulletin. 71. https://doi.org/10.1016/j.marpolbul.2013.03.024

Cordier, T., Forster, D., Dufresne, Y., Martins, CIM., Stoeck, T., Pawlowski, J. (2018) Supervised machine learning outperforms taxonomy-based DNA metabarcoding applied to biomonitoring. Molecular Ecology Resources. 18, 6. https://doi.org/10.1111/1755-0998.12926

Coull, B. C., Chandler, G. T. (2001). Benthos (Meiobenthos). In: Steele, J. H., Turekian, K., Thorpe, S. A. (Eds). Encyclopedia of Ocean Sciences. 1705-1711.

Drummond, A. J., Newcomb, R. D., Buckley, T. R., Xie, D., Dopheide, A., Potter, B. C., … Nelson, N. (2015). Evaluating a multigene environmental DNA approach for biodiversity assessment. Gigascience. 4, 46. https://doi.org//10.1186/s13742-015-0086-1

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 26 (19), 2460-2461. https://doi.org/10.1093/bioinformatics/btq461

Elbrecht, V., Peinert, B. & Leese, F. (2017). Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. Ecology and Evolution. 7, 6918-6926. https://doi.org/10.1002/ece3.3192

Elbrecht, V. & Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative metabarcoding protocol. Plos One. 10 (7), e0130324. https://doi.org/10.1371/journal.pone.0130324

Environmental Agency. (2016). Water Framework Directive assessment: estuarine and coastal waters. Retrieved from https://www.gov.uk/guidance/water-framework-directive-assessment-estuarine-and-coastal-waters

Froeslev, T. G., Kjoeller, R., Bruun, H. H., …Hansen, A. J. (2017). Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nature Communications. 8, 1188. https://doi.org/10.1038/s41467-017-01312-x

Geller, J., Meyer, C., Parker, M. & Hawk, H. (2013). Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application

485 in all-taxa biotic surveys. Molecular Ecology Resources. 13 (5), 851-861.
486 https://doi.org/10.1111/1755-0998
487 Glassman, S., Martiny, J. (2018). Broadscale patterns are robust to use of exact sequence
488 variants versus operational taxonomic units. mSphere. 3 (4), e00148-18.
489 https://doi.org/10.1128/mSphere.00148-18
490 Haenel, Q., Holovachov, O., Jondelius, U., Sundberg, P., Bourlat, S. J. (2017). NGS-based
491 biodiversity and community structure analysis of meiofaunal eukaryotes in shell sand
492 from Hållö island, Smögen, and soft mud from Gullmarn Fjord, Sweden. Biodiversity
493 Data Journal. 5, e12731. https://doi.org/10.3897/BDJ.5.e12731
494 Hautier, Y., Tilman, D., Isbell, F., Seabloom, E. W., Borer, E. T. & Reich, P. B. (2015) Plant
495 ecology. Anthropogenic environmental changes affect ecosystem stability via
496 biodiversity. Science. 348, 336-340. https://doi.org/10.1126/science.aaa1788
497 Hirai, J., Kuriyama, M., Ichikawa, T., Hidaka, K. & Tsuda, A. (2015). A metagenetic
498 approach for revealing community structure of marine planktonic copepods. Molecular
499 Ecology Resources. 15 (1), 68-80. https://doi.org/ 10.1111/1755-0998
500 Illumina. 16S Metagenomic sequencing library preparation. In:
501 http://support.illumina.com/content/dam/illumina-
502 support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-
503 library-prep-guide-15044223-b.pdf. Accessed date: 13 June 2015.
504 Keck, F., Vasselon, V., Tapolczai, K., Rimet, F. & Bouchez, A. (2017). Freshwater
505 biomonitoring in the information age. Frontiers in Ecology and the Environment. 15
506 (5), 266-274. https://doi.org/10.1002/fee.1490
507 Kelly, R. P., Closek, C. J., O'Donnell, J. L., Kralj, J. E., Shelton, A. O. & Samhouri, J. F.
508 (2017). Genetic and manual survey methods yield different and complementary views
509 of an ecosystem. Frontiers in Marine Science. 3, 283.
510 https://doi.org/10.3389/fmars.2016.00283
511 Kvist, S. (2013). Barcoding in the dark? A critical view of the sufficiency of zoological DNA
512 barcoding databases and a plea for broader integration of taxonomic knowledge.
513 Molecular Phylogenetics and Evolution. 69 (1), 39.
514 https://doi.org/10.1016/j.ympev.2013.05.012
515 Lanzén, A., Lekang, K., Jonassen, I., Thompson, E. M. & Troedsson, C. (2016). High-
516 throughput metabarcoding of eukaryotic diversity for environmental monitoring of
517 offshore oil- drilling activities. Molecular Ecology. 25 (17), 4392-4406.
518 https://doi.org/10.1111/mec.13761

519 Laroche, O., Wood, S. A., Tremblay, L. A., Lear, G., Ellis, J. I., Pochon, X. & Reimer, J.
520 (2017). Metabarcoding monitoring analysis: the pros and cons of using co-extracted
521 environmental DNA and RNA data to assess offshore oil production impacts on benthic
522 communities. Peerj. 5, e3347. https://doi.org/10.7717/peerj.3347
523 Lejzerowicz, F., Esling, P., Pillet, L., Wilding, T. A., Black, K. D. & Pawlowski, J. (2015).
524 High-throughput sequencing and morphology perform equally well for benthic
525 monitoring of marine ecosystems. Scientific Reports. 5, 13932. https://doi.org/
526 10.1038/srep13932
527 Leray, M., Meyer, C. P. & Mills, S. C. (2015). Metabarcoding dietary analysis of coral
528 dwelling predatory fish demonstrates the minor contribution of coral mutualists to their
529 highly partitioned, generalist diet. Peerj. 3, e1047. https://doi.org/ 10.7717/peerj.1047
530 Leray, M. & Knowlton, N. (2015). DNA barcoding and metabarcoding of standardized
531 samples reveal patterns of marine benthic diversity. Proceedings of the National
532 Academy of Sciences of the United States of America. 112 (7), 2076-2081.
533 https://doi.org/ 10.1073/pnas.1424997112
534 Leray, M., Ho, S. L., Lin, I. J., Machida, R. J. (2018). MIDORI server: a webserver for
535 taxonomic assignment of unknown metazoan mitochondrial-encoded sequences using a
536 curated database. Bioinformatics. 34 (21), 3753-3754.
537 https://doi.org/10.1093/bioinformatics/bty454
538 Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., … Machida, R.
539 J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial
540 COI region for metabarcoding metazoan diversity: application for characterizing coral
541 reef fish gut contents. Frontiers in Zoology. 10, 34. https://doi.org/10.1186/1742-9994-
542 10-34
543 Lobo, J., Shokralla, S., Costa, M. H., Hajibabaei, M., Costa, F, O. (2017). DNA
544 metabarcoding for high-throughput monitoring of estuarine microbenthic communities.
545 Scientific Reports. 7, 15618. https://doi.org/ 10.1038/s41598-017-15823-6
546 Morad, Y. T., Dubinsky, Z., Iluz, D. (2017) Meiobenthos assemblages as bioindicators for
547 coastal pollution assessment. Open Journal of Marine Science. 7, 409-423.
548 https://doi.org/10.4236/ojms.2107.73028
549 Needham, M. D., Sachdeva, R., Furham, A. J. (2017). Ecological dynamics and co-
550 occurrence among marine phytoplankton, bacteria and myoviruses shows
551 microdiversity matters. The ISME Jounral. 11, 1614-1629.
552 https://doi.org/10.1038/ismej.2017.29

553 Pearman, J. & Irigoien, X. (2015). Assessment of zooplankton community composition along
554 a depth profile in the central Red Sea. Plos One. 10 (7), e0133487.
555 https://doi.org/10.1371/journal.pone.0133487
556 Pearman, J. K., Leray, M., Villalobos, R., Mchida, R. J., Berumen, M. L., Knowlton, N.,
557 Carvalho S. (2018). Cross-shelf investigation of coral reef cryptic benthic organisms
558 reveals diversity patterns of the hidden majority. Scientific Report. 8, 8090.
559 http://doi.org/ 10.1038/s41598-018-26332-5
560 Poikane, S., Johnson, R. K., Sandin, L., Schartau, A. K., Solimini, A. G., Urbanič, G., …
561 Böhmer, J. (2016). Benthic macroinvertebrates in lake ecological assessment: A review
562 of methods, intercalibration and practical recommendations. The Science of the Total
563 Environment. 543, 123. https://doi.org/10.1016/j.scitotenv.2015.11.021
564 Porter, T. M., Hajibabaei, M. (2018). Scaling up: A guide to high-throughput genomic
565 approaches for biodiversity analysis. Molecular Ecology. 2, 313-338. https://doi.org/
566 10.1111/mec.14478
567 Ransome, E., Geller, J. B., Timmer, M., Leray, M., Mahardini, A., Sembiring, A., Collins, A.
568 G., Meyer, C. (2017). The importance of standardization for biodiversity comparisons;
569 A case study using autonomous reef monitoring structures (ARMS) and metabarcoding
570 to measure cryptic diversity on Mo'orea coral reefs, French Polynesia. PLOS ONE. 12,
571 e0175066. https://doi.org/10.1371/journal.pone.017066
572 Ratnasingham, S. & Hebert, P. D. N. (2007). bold: The Barcode of Life Data System.
573 Molecular Ecology Notes. 7 (3), 355-364. https://doi.org/ 10.1111/j.1471-
574 8286.2007.01678.x
575 Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahe, F. (2016). VSEARCH : a versatile
576 open source tool for metagenomics. PeerJ. 4, e2584. https://doi.org/10.7717/peerJ.2584
577 Roussel, J., Paillisson, J., Tréguier, A. & Petit, E. (2015). The downside of eDNA as a survey
578 tool in water bodies. Journal of Applied Ecology. 52 (4), 823-826.
579 https://doi.org/10.1111/1365-2664.12428
580 Stoeckle, M., Soboleva, L. & Charlop-Powers, Z. (2017). Aquatic environmental DNA
581 detects seasonal fish abundance and habitat preference in an urban estuary. PLoS One.
582 12 (4), e0175186. https://doi.org/10.1371/journal.pone.0175186
583 Taberlet, P., Coissac, E., Hajibabaei, M. & Rieseberg, L. H. (2012). Environmental DNA.
584 Molecular Ecology. 21, 1789-1793. https://doi.org/10.1111/j.1365-294X.2012.05542.x

585 Thomsen, P. F. & Willerslev, E. (2015). Environmental DNA - an emerging tool in
586 conservation for monitoring past and present biodiversity. Biological Conservation.
587 183, 4-18. https://doi.org/10.1016/j.biocon.2014.11.019
588 Torti, A., Lever, M. A. & Jørgensen, B. B. (2015). Origin, dynamics, and implications of
589 extracellular DNA pools in marine sediments. Marine Genomics. 3, 185-196.
590 https://doi.org/10.1016/j.margen.2015.08.007
591 Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., … Dejean, T.
592 (2016). Next-generation monitoring of aquatic biodiversity using environmental DNA
593 metabarcoding. Molecular Ecology. 25 (4), 929-942. https://doi.org/
594 10.1111/mec.13428
595 Wangensteen, O. S., Palacin, C., Guardiola, M., Turon, X. (2018). DNA metabarcoding of
596 littoral hard-bottom communities: high density and database gaps revealed by two
597 molecular markers. PeerJ. 6, e4705. https://doi.org/ 10.7717/peerj.4705
598 Wilcox, M. T., McKelvey, S. K., Young, M. K., Jane, F. S., Lowe, H. W., Whiteley, R. A. &
599 Schwartz, K. Michael. (2013). Robust detection of rare species using environmental
600 DNA: the importance of primer specificity. Plos One. 8 (3), e59520.
601 https://doi.org/10.1371/journal.pone.0059520
602 Worsfold, T.M., Hall, D. (2010). Guidelines for processing marine microbenthic invertebrate
603 samples: a Processing Requirements Protocol (Version 1.0, June 2010). National
604 Marine Biological Analytical Quality Control Scheme. Available online at
605 http://www.nmbaqcs.org/media/1175/nmbaqc-inv-prp-v10-june2010.pdf
606 Yamamoto, S., Masuda, R., Sato, Y., Sado, T., Araki, H., Kondoh, M., Minamoto, T. &
607 Miya, M. (2017). Environmental DNA metabarcoding reveals local fish communities in
608 a species-rich coastal sea. Scientific Reports. 12 (4), e0176608. https://doi.org/
609 10.1038/srep40368
610 Zimmermann, J., Glöckner, G., Jahn, R., Enke, N. & Gemeinholzer, B. (2015).
611 Metabarcoding vs. morphological identification to assess diatom diversity in
612 environmental studies. Molecular Ecology Resources. 15 (3), 526-542.
613 https://doi.org/10.1111/1755-0998
614
615
616
617
618
619

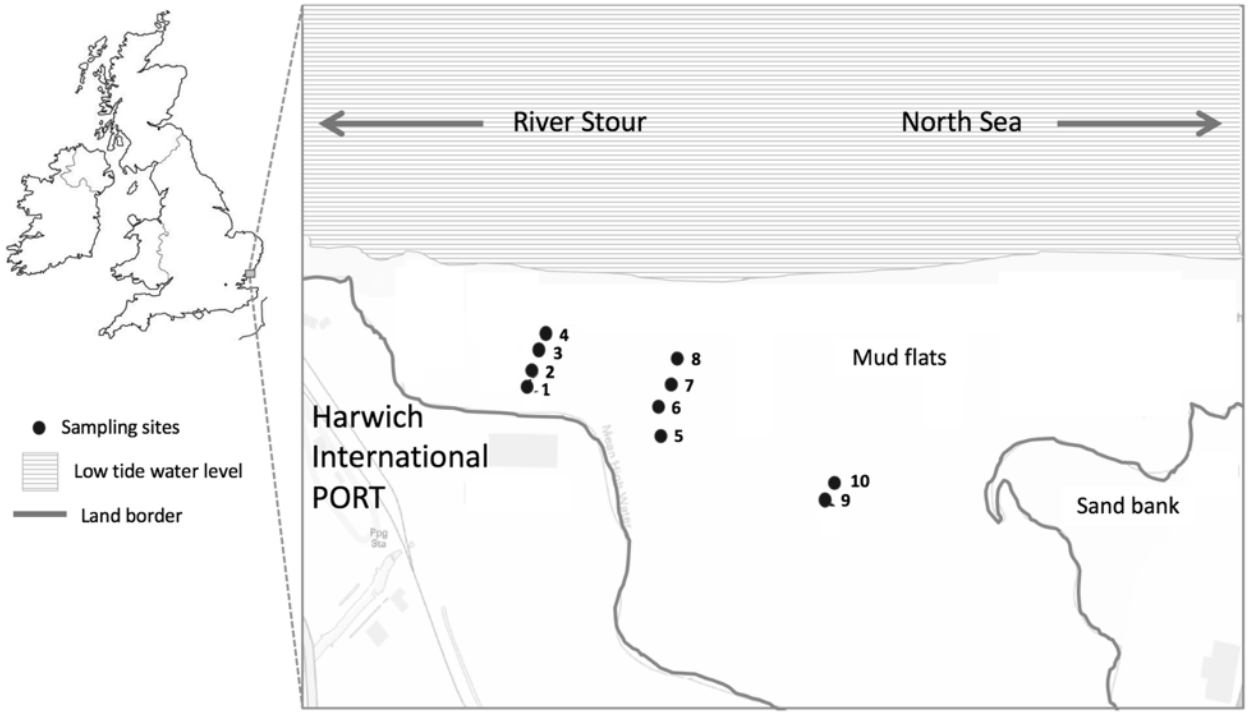
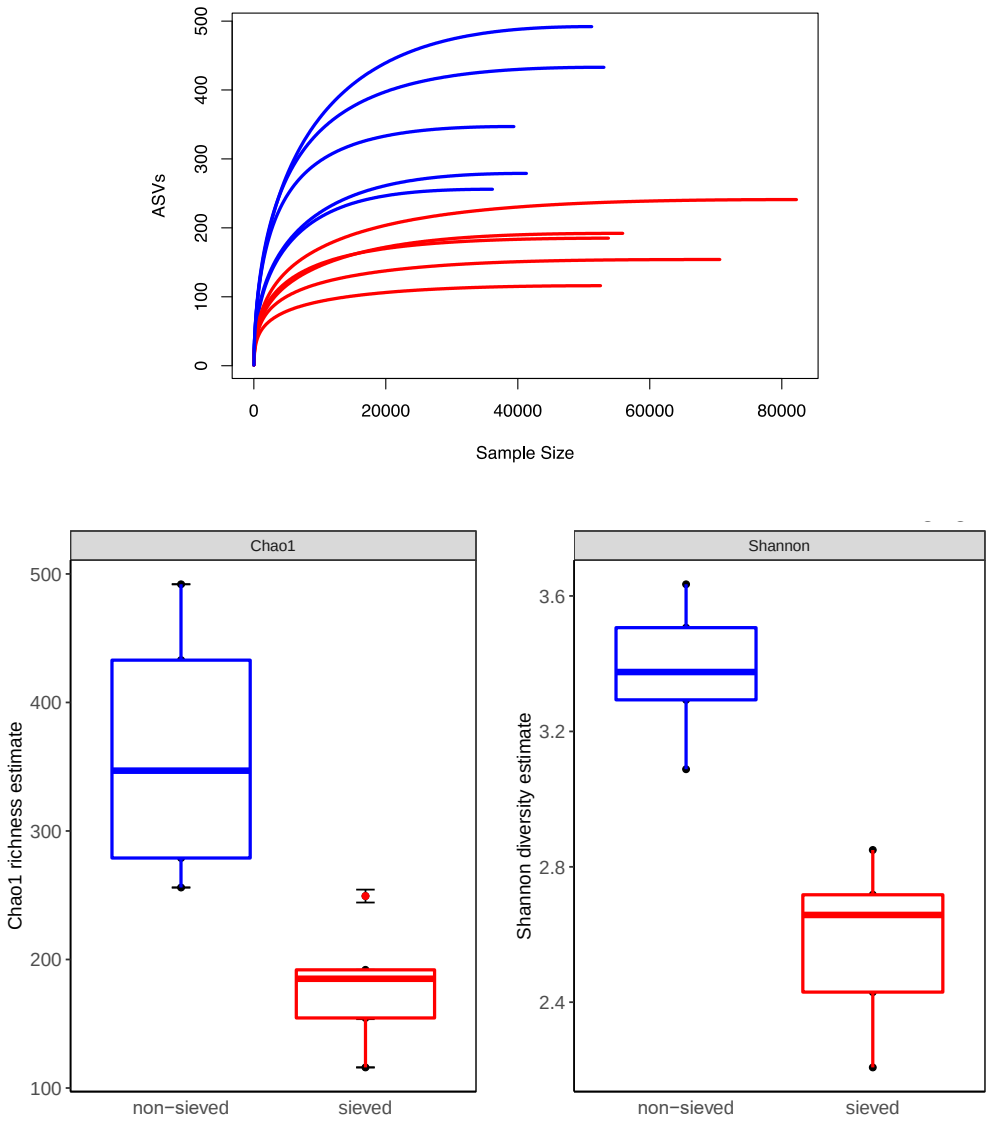


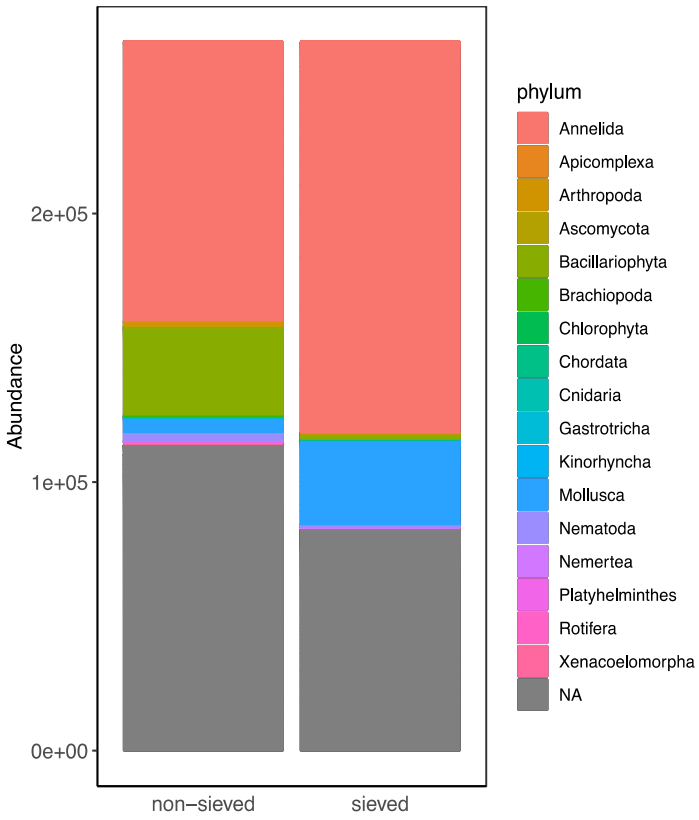
Figure 1. Map showing the location of extracted benthic cores at the Harwich International Port estuary, Essex, U.K. GPS coordinates for all ten sample sites can be found in Table S1.

643 **Table 1.** Summary of the number of raw reads, number of reads post filtering, merging and
644 chimera removal steps in the DADA2 pipeline and the number of ASVs prior to and post LULU
645 curation across sieved and un-sieved samples.

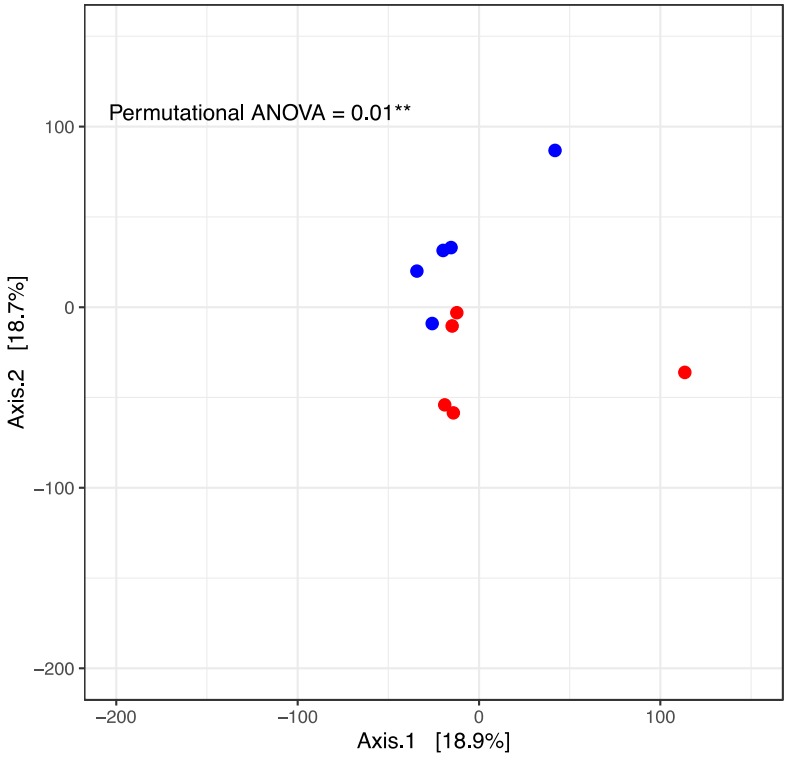
Sample	Processing step	Raw reads	Number of reads post DADA2 filtering step	Number of reads post DADA2 merging step	Number of reads post DADA2 chimera removal step	Number of ASVs prior to LULU curation step	Number of ASVs post LULU curation step
1	Sieved	1,053,425	82,449	80,381	77,456	208	154
2	Not sieved	1,135,799	58,899	57,842	57,247	349	279
3	Sieved	1,034,144	70,674	69,324	67,298	143	116
4	Not sieved	1,131,688	48,468	47,114	46,476	432	347
5	Sieved	1,1129,646	61,482	59,960	58,199	231	185
6	Not sieved	870,689	47,965	46,003	43,464	326	257
7	Sieved	1,175,309	94,870	93,408	91,223	313	241
8	Not sieved	1,258,270	68,755	66,637	65,155	547	433
9	Sieved	1,119,207	60,351	59,293	58,359	225	192
10	Not sieved	1,165,277	56,958	56,958	56,293	620	492



662 **Figure 2.** a) Rarefaction curves of ASV diversity in sieved (red) and non-sieved (blue)
663 samples and b) boxplot of Chao1 and Shannon estimates of ASV richness and diversity in
664 sieved (red) and non-sieved (blue) samples.



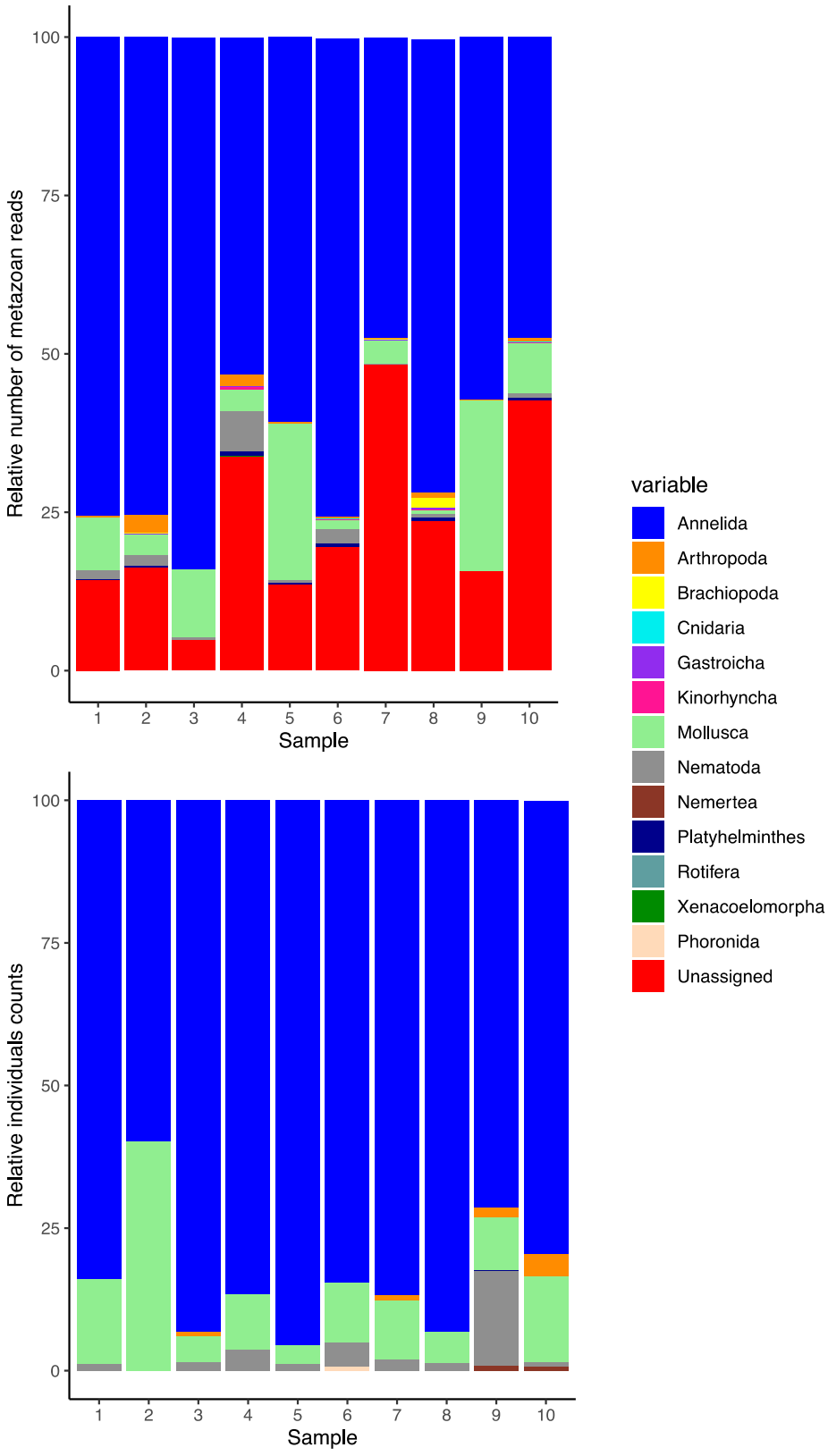
665 **Figure 3.** Barplot comparing relative abundances of normalised numbers of reads per phyla
666 across non-sieved and sieved samples.



667 **Figure 4.** Principle Coordinates Analysis (PCoA) plot of ASV abundances across sieved (red)
668 and unsieved (blue) samples, using Euclidean distances. The result of a permutational ANOVA
669 test comparing the communities of sieved and unsieved samples is displayed within the plot.

Table 2. Summary table of the number of marine species, per phylum, identified across all metabarcoding and morphological samples, as well as in un-sieved and sieved metabarcoding samples. Cells with dashes indicate the phyla in question were not targeted by the morphological approach. Cells with species count as 0 indicate that whilst no species was identified, the presence of this phylum was detected via the identification of an ASV to a higher taxonomic rank (e.g. class, order or family level).

Phylum	Number of species detected across all metabarcoding samples	Number of species identified across all morphological samples	Number of species in sieved metabarcoding samples	Number of species in un-sieved metabarcoding samples
Annelida	11	16	9	11
Apicomplexa	0	-	0	0
Arthropoda	1	1	0	1
Bacillariophyta	0	-	0	0
Bryozoa	0	-	0	0
Chlorophyta	1	-	1	1
Chordata	1	-	0	1
Cnidaria	1	-	1	1
Echinodermata	0	-	0	0
Gastrotricha	0	-	0	0
Kinorhyncha	1	-	0	1
Mollusca	6	8	5	5
Nematoda	1	0	1	1
Nemertea	0	0	0	0
Platyhelminthes	1	0	1	0
Rotifera	0	-	0	0
Xenacoelomorpha	0	-	0	0
Total number of sp.	24	25	18	22



676 **Figure 5**. Abundance barplots displaying the composition of organisms identified at phylum
677 level in both the morphological and metabarcoding analyses. Colour legend applies to both
678 graphs.

Table 3. Summary table of the taxa identified in the morphological analysis and if those taxa were i) identified by the metabarcoding approach, ii) if not, what the closest taxonomic level to that taxa is, iii) whether the taxa is represented in the reference dataset used in this study and finally iv) whether the taxa in question is detected using BLASTn, a different taxonomy assignment tool.

Taxa detected by morphological approach	Is this taxon detected by the metabarcoding approach using the RDP classifier dataset?	If not detected, what is closest taxonomic level to this taxon that is detected?	Is this taxon represented in the MIDORI-UNIQUE reference dataset?	Is this taxon detected using BLASTn and the NCBI nucleotide collection reference database?
Pholoe	No	Phyllodocida order	Yes	No
Phyllodoce mucosa	No	*Phyllodoce* genus	Yes	Yes
Eteone (Type 1)	No	Phyllodocidae family	Yes	Yes
Glycera tridactyla	No	Phyllodocidae family	Yes	No
Exogone naidina	No	Phyllodocidae family	No (only genus present)	No
Sphaerosyllis tetralix	No	Phyllodocida order	No (only genus present)	No
Hediste diversicolor	Yes	-	Yes	Yes
Nephtys hombergii	Yes	-	Yes	Yes
Pygospio elegans	No	Spionidae family	Yes	Yes
Streblospio	No	Spionidae family	Yes	Yes
Cirratulidae	Yes	-	Yes	Yes
Aphelochaeta	No	Cirratulidae family	Yes	No
Cirriformia tentaculata	No	Cirratulidae family	Yes	Yes
Tharyx	No	Terebellida family	No (only family present)	No
Tharyx robustus	No	Terebellida family	No (only family present)	No
Tharyx killariensis	No	Terebellida family	No (only family present)	No
Cossura pygodactylata	No	No close taxonomic level detected	No (only genus present)	No
Capitella	Yes	-	Yes	No
Galathowenia	No	No close taxonomic level detected	Yes	No
Melinna palmata	No	Ampharetidae family	n (only genus present)	No
Manayunkia	No	Spionidae family	Yes	No
Tubificoides amplivasatus	No	*Tubificoides* genus	Yes	Yes
Tubificoides benedii	Yes	-	Yes	Yes
Tubificoides pseudogaster (agg.)	Yes	-	Yes	Yes
Eusarsiella zostericola	No	Myodocopida order	No (only genus present)	No
Peringia ulvae	Yes	-	Yes	Yes
Limapontia depressa	No	Limapontiidae family	Yes	Yes

Nuculidae	Yes	-	Yes	Yes
Nucula nitidosa	No	*Nuxcula* genus	Yes	Yes
Mytilidae	No	Bivalvia class	Yes	No
Cardiidae	Yes	-	Yes	Yes
Cerastoderma edule	Yes	-	Yes	Yes
Limecola balthica	No	*Limecola* genus	Yes	Yes
Abra tenuis	No	Cardiida order	No (only genus present)	No
Scrobicularia plana	Yes	-	Yes	Yes
Phoronis	No	Phoroniformea sub-phylum	Yes	No