The detection of novel and resident marine species using environmental DNA metabarcoding of sediment and water

Mr. Luke E. Holman[1]* (ORCID: 0000-0002-8139-3760)

Dr. Mark de Bruyn[2,3]

Prof. Simon Creer[2]

Prof. Gary Carvalho[2]

Dr. Julie Robidart[4]

Dr. Marc Rius[1,5] (ORCID: 0000-0002-2195-6605)

[1] School of Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton, United Kingdom

[2] Molecular Ecology and Fisheries Genetics Laboratory, School of Natural Sciences, Bangor University, United Kingdom

[3] The University of Sydney, School of Life and Environmental Sciences, Australia

[4] Ocean Technology and Engineering Group, National Oceanography Centre Southampton, United Kingdom

[5] Centre for Ecological Genomics and Wildlife Conservation, University of Johannesburg, South Africa

*Corresponding Author

Email: l.e.holman@soton.ac.uk

Abstract

Biodiversity monitoring is often challenging as it requires taxonomic expertise and is difficult to automate. A non-invasive method that tackles these challenges and detects organisms that are often overlooked by existing monitoring techniques is the use of environmental DNA (eDNA). Despite considerable progress in recent years, studies evaluating the effectiveness of eDNA surveys for the detection of novel and resident species remain rare, and little is known about the effects that different environmental sample types have on species detectability. We evaluated the ability of eDNA metabarcoding to detect marine species in urban coastal environments, and compared the results with previously published biodiversity surveys. We collected sediment and water samples and performed eDNA metabarcoding of 18S rRNA and COI genes. We detected vastly different species assemblages from the two environmental sample types; with sediment containing in some cases more than twice the number of operational taxonomic units than water. In addition, species detection within phyla was highly variable between sediment and water samples from the same sampling site. Interestingly, we found as much variation in assemblage diversity among environmental sample types as amongst geographically-segregated sampling sites. Finally, eDNA metabarcoding detected three previously unreported non-indigenous species (NIS), suggesting that the use of eDNA surveys can significantly reduce the time between introduction and detection of NIS and increase the likelihood of successful control / eradication. We conclude that careful consideration of type of environmental sample is needed when conducting biodiversity monitoring using eDNA, especially for surveys aiming to detect community change and NIS.

Keywords

Aquatic, biomonitoring, conservation genetics, DNA metabarcoding, HTS, non-native, NGS.

1. Introduction

Anthropogenic activities are causing a global decrease in biodiversity (Sala and Knowlton 2006; Butchart et al. 2010) that negatively affects ecosystem services and function (Worm et al. 2006). Such impacts create an urgent need for conservation tools that rapidly and accurately monitor species diversity. Biodiversity surveys have been criticised for their lack of standardisation and taxonomic resolution (Oliver & Beattie 1993; Fitzpatrick *et al.* 2009). One approach that has the potential to overcome these limitations is the use of nucleic acids found in environmental samples, such as water, soil or sediment, to infer presence or absence of living organisms in the local environment (Thomsen & Willerslev, 2014). This genetic material, known as environmental DNA (hereafter eDNA), is a poly-disperse mixture of tissue, cells, subcellular fragments and extracellular DNA lost to the environment by organisms (Turner *et al.* 2014; Sassoubre *et al.* 2016). Studies using eDNA focus on targeted detection (single species) methods such as qPCR (Dougherty *et al.* 2016; Simpson *et al.* 2017; Wood *et al.* 2017; Kim *et al.* 2018), or community (multi-species) methods such as metabarcoding (Borrell *et al.* 2017; Grey *et al.* 2018; Lacoursière-Roussel *et al.* 2018). Samples can be collected with minimal training and once the methodology is optimised, surveys are highly amenable to automation (McQuillan & Robidart 2017). Thus, eDNA surveys are highly informative (but see considerations for ensuring validity and replicability in Goldberg *et al.* 2016) and can complement other biodiversity monitoring methods (Deiner *et al.* 2017). Recent work has identified a vast range of viable protocols for the collection, extraction and detection of target nucleic acids from different environmental samples (Deiner et al., 2018; Spens *et al.* 2016; Sellers *et al.* 2018). Despite this, little is known about the effects that different environmental sample types have on species detectability using eDNA (Shaw *et al.* 2016).

Biodiversity monitoring is normally performed by conducting surveys standardised by time and by reaching a species discovery asymptote (Ashton *et al.* 2006; Campbell, Gould & Hewitt 2007; Bishop *et al.* 2015). The results of such surveys often restrict to species that are being targeted

at the time, with no ability to retrospectively separate erroneously grouped species in light of new discoveries. This is a critically important aspect for biodiversity monitoring in the sea, mainly because between 9,000-35,000 marine species (2.7% of the total number of estimated marine species) are considered cryptic (i.e., morphologically similar but genetically distinct) (Appeltans *et al.* 2012). Indeed, many widespread sessile marine species contain cryptic lineages as revealed by genetic studies, highlighting the need for an integrated morphological and genetic approach to accurately detect these species (Pérez-Portela *et al.* 2013; Rius *et al.* 2017).

Non-indigenous species (NIS) are those that have been transported from their native range through human action into a novel geographic location. The impacts of NIS pose a severe threat to agriculture, human health and the conservation of ecosystems and biodiversity (Bax *et al.* 2003; Lovell, Stone & Fernandez 2006; Ricciardi *et al.* 2013; Mazza *et al.* 2014). Most marine NIS have spread globally via vectors such as transoceanic shipping or the construction of canals interconnecting large water bodies (Molnar *et al.* 2008; Nunes *et al.* 2014). At fine (10s of km) geographical scales, other vectors such as intraregional boating significantly enhance the spread and subsequent impact of NIS (Clarke Murray *et al*. 2011). Along coastal areas, studies have highlighted the importance of monitoring marinas and harbours (Ashton *et al.* 2006), as these are hotspots of NIS and together with other marine infrastructure (e.g. breakwaters, artificial reefs) promote the spread of NIS (Airoldi *et al.* 2015). In these habitats, NIS often outcompete native species and dominate artificial hard substrata (Glasby *et al.* 2007; Dafforn *et al*. 2009). Marinas and harbours have distinct ecological and physico-chemical conditions compared to the surrounding natural environment (Rivero *et al.* 2013; Foster *et al.* 2016). Consequently, there is a need for specific sampling and surveying protocols to study both native and NIS in these ecologically distinct environments.

Here we used eDNA metabarcoding to examine how species detection differs between environmental sample types collected in several distinct marinas. We first documented the differences in both alpha and beta diversity from sediment and water samples. We then

compared the eDNA metabarcoding results with previously published biodiversity data to identify key NIS that are both introduced in the study region and / or elsewhere. Subsequently, we identified a number of previously unrecorded NIS in the study region. Finally we discuss the strengths and weaknesses of eDNA metabarcoding for detecting marine NIS, and how this technique can help conservation efforts for both preserving native biodiversity and mitigating the deleterious effects of NIS.

2. Methods

2.1 Study sites

Four marinas were selected from around the United Kingdom to represent variation in modelled invasion potential (Pearce, Peeler & Stebbing 2012), known NIS present (Bishop *et al.* 2015) and surrounding benthic habitat type (Calewaert *et al.* 2016). Importantly, all chosen marinas have been surveyed previously for NIS and so there is a good understanding of the expected NIS in these areas (Wood, Bishop & Yunnie 2015a; b; Wood *et al.* 2016). Marina access was contingent on anonymity and so marina names and exact locations are not presented, Fig.1a shows approximate locations. Marina *TQ* is an open marina subject to tides and varying salinity located in Southampton Water on the north coast of the English Channel. Marina *PQ* is a loch marina open during high tide to the Bristol Channel and the Celtic Sea. Marina *TB* is located at the mouth of the River Blackwater open to the North Sea. Marina *HH* is located on the Isle of Anglesey and is open to the Celtic Sea.

2.2. Environmental DNA sampling

Surveys were conducted during May 2017. A total of 24 sampling points were randomly selected within each site. At each point 50ml of water was collected from 10cm below the surface using a sterile 60ml Luer lock syringe and filtered through a 0.22mm polyethersulfone Sterivex filter

(Merck Millipore, Massachusetts USA). After collecting seawater from eight locations (400ml total volume) the filter was changed, resulting in a total of three filters per site. To test the effect of different sample preservation methods sampling was performed in duplicate. One set of three filters had ~1.5ml sterile Longmire's solution (100mM Tris,10mM EDTA, 10mM NaCl, 0.5% SDS) applied in the inlet valve (Renshaw et al. 2015). The second set of three filters were kept on ice for no longer than eight hours before being frozen at -20°C. During the surveys, a sediment sample was collected at the first water sampling site and then after every 3rd water sample, for a total of 9 per site. We used a UWITEC Corer (UWITEC, Mondsee, Austria) to collect a sediment core (600mm tall x 60mm diameter). Using a sterile disposable spatula, a subsample of 10-20g of sediment was taken from the top 2cm of the core, taking care to avoid sampling the sides of the core. The subsample was stored in a sterile plastic zip container and kept on ice for no longer than eight hours before begin frozen at -80°C. Due to equipment malfunction no sediment sample could be taken for Site HH. Disposable gloves were changed after collection of each sample. All reused equipment was washed thoroughly and soaked in 10% bleach between sites, before rinsing in DNAse-free sterile water.

2.3. eDNA extraction

DNA extractions were performed in a PCR–free cleanroom, separate from main laboratory facilities. No high copy templates, cultures or amplicons were permitted in this sterile laboratory. DNA extractions from water samples followed Spens et al. (2016) using the $SX_{CAPSULE}$ method. Briefly, preservative solution was removed from the outlet and filters were dried at room temperature for two hours, 720µl Qiagen buffer ATL (Qiagen, Hilden, Germany) and 80µl Proteinase K was added to the filter and all samples were digested overnight at 56ºC. After digestion, samples were processed using the Qiagen DNeasy Blood and Tissue Kit as per manufacturer instructions. The final elution was 200µl PCR grade water.

Sediment extractions were performed using the Qiagen DNeasy Powermax Soil Kit using the manufacturer recommended protocol. For each site the nine samples were mixed to form three pooled samples; for each extraction, 10g of pooled sample was processed. A total of ten samples were processed, three each per site with a single no blank, as per manufacturer's instructions.

2.4. Inhibition testing

To ensure extracted DNA was free of PCR inhibitors a Primer Design Real-Time PCR Internal Control Kit (PrimerDesign, Southampton, UK) was used. qPCR reactions were performed following the manufacturer's protocol. Inhibition due to co-purified compounds from DNA extraction protocols would produce an increase in cycle threshold number in comparison to no template controls. All samples were successfully processed and no samples showed indication of PCR inhibition.

2.5. Primer selection & library preparation

Two sets of primers were chosen for metabarcoding the environmental samples: a 313bp section of the standard DNA barcoding region of the cytochrome c oxidase subunit I gene (COI) using primers described in Leray et al. (2013); and a variable length target of the hypervariable v4 region of the nuclear small subunit ribosomal DNA (18S) using primers from Zhan *et al.* (2013). Sequencing libraries were prepared using a 2-step PCR approach as detailed in Bista *et al*. (2017), this method amplifies the target region in PCR 1 annealing universal adapters onto which sample specific indices and sequencing primers are annealed in PCR 2. In contrast to Bista *et al*. (2017) we used unique dual-matched indexes for PCR 2 to avoid index crosstalk associated with combinatorial indexing (MacConaill et al. 2018). PCR 1 was prepared in a PCR–free room, separate from main laboratory facilities. PCR conditions and reaction volumes are detailed in Appendix A.1. Blank filters, DNA extraction kits and positive controls where collected,

extracted and sequenced as the experimental treatments (detailed in Appendix A.2). Samples were pooled at an equimolar ratio and sequenced using the Illumina MiSeq instrument (Illumina, San Diego, USA) with a V3 2 x 300bp kit.

2.6. Bioinformatic analyses

Samples were demultiplexed using the Illumina MiSeq control software (v 2.6.2.1). The demultiplexed data was analysed using a custom pipeline written in the R programming language (R Core Team 2018) hosted at https://github.com/leholman/metabarTOAD, the steps are as follows. Forward and reverse paired end reads were merged using the -fastq_mergepairs option of USEARCH v.10.0.240 (Edgar 2013) with maximum difference of 15, percent identity of 80% and quality filter set at maximum expected errors of 1. Both the forward and reverse primer sequences were matched using Cutadapt v.1.16 (Martin 2011) and only sequences containing both primer regions were retained. Sequences were discarded if they were outside of a defined length boundary (303-323bp for COI, 375-450bp for 18S) using Cutadapt. Sequences were then pooled, singletons were discarded and sequences were quality filtered with a maximum expected error of 1 using the -fastq_filter option of vsearch v.2.4.3 (Rognes *et al.* 2016). Sequences were then denoised and chimeras filtered using the unoise3 algorithm implemented in USEARCH. The resultant operational taxonomic units (OTUs) were curated using the LULU package v.0.1.0 in R (Frøslev *et al.* 2017). An OTU by sample table was produced by mapping the merged and trimmed reads against the curated OTUs using USEARCH, with the raw query read assigned to the OTU with the best match (highest e value) within 97% identity. The OTU x sample table was filtered in R as follows. To minimise the chance of spurious OTUs being included in the final dataset any record with less than 3 raw reads were changed to zero and any OTU that did not appear in more than one sample was removed from the analysis. OTUs found in negative controls were removed from the analysis.

2.7. Taxonomic assignment

Assigning correct taxonomy to an unknown set of marine sequences can be challenging as large databases require vast computational resources for query matching; many databases contain errors and the taxonomy of some marine groups is uncertain. With such limitations in mind, we assigned taxonomy using a BLAST v.2.6.0+ search (Camacho *et al.* 2009) returning the single best hit (largest e value) from databases within 97% of the query using a custom R script to parse the raw blast results. The MIDORI database (UNIQUE_20180221) (Machida *et al.* 2017) was used for the COI data and the SILVA database (SSU r132) (Quast *et al.* 2013) was used for the 18S rRNA data. The match taxa tool from the World Register of Marine Species (WoRMS) (WoRMS Editorial Board 2018) was used to filter the data for marine species and check the classification. Remaining annotations were checked against the World Register of Introduced Marine Species (WRIMS) (Ahyong *et al.* 2018) to determine non-indigenous status.

2.8. Statistical analyses

All statistical analyses were conducted in R v3.5.0. The Vegan R package v.2.5.2 (Oksanen *et al.* 2011) was used to rarefy samples to the minimum sample read depth for each amplicon. The number of OTUs per site/condition was calculated as the number of OTUs with a non-zero number of normalized reads after summing the reads across all three site level replicates. To test if there was a significant difference between the number of OTUs generated by sediment and water eDNA, individual non-summed replicate sample data was used to build a two-way ANOVA model with the formula *number_of_OTUs~sedimentorwater*site* implemented in R using the function aov(). Non-metric multidimensional scaling ordination plots were generated from Bray-Curtis dissimilarity values derived using Vegan. A Permutation Analysis of Variance (PERMANOVA) (Balakrishnan et al. 2014) was performed using the Bray Curtis dissimilarity following the model *dissimilarity_matrix~sedimentorwater*site* implemented in R using the

function *adonis* from the *vegan* package. OTUs with taxonomic assignment were separated into those found in sediment, water or both media and the OTUs were then collapsed at the Phylum level to explore taxonomic patterns of detection in water or sediment. Phyla with less than eight OTUs were combined. To test for non-random counts of species detection between water and sediment within taxa an exact binomial test was performed between counts of species detected in water and sediment. Half the number of counts for species detected in both water and sediment were added to water and sediment, with non-integer values conservatively rounded down to the nearest whole number. A Bonferroni correction for multiple comparisons was applied across the p values from the exact binomial tests. Records from manual surveys previously conducted for non-native invertebrates at the sample sites (Wood, Bishop & Yunnie 2015a; b; Wood *et al.* 2016) were compared with the detected species from metabarcoding data.

3. Results

3.1. Raw sequencing results & OTU generation

Sequencing produced a total of 17.8 million paired end reads, with 15.2 million sequences remaining after paired end read merging and quality filtering. The average number of sequences per sample after filtering (excluding control samples) was 200,185 ± 64,019 (s.d). No template control samples contained an average of 811 ± 3,402 (s.d) sequences. One control sample contained ~15,000 sequences that mapped to an operational taxonomic unit (OTU) that had 100% identity match to a sequence of a terrestrial fungi (Genbank: FJ804151.1), excluding this OTU gives an average of 51 ± 94 (s.d) sequences per no-template control sample. Denoising produced 8,069 OTUs for COI and 2,433 for 18S with 6,435 and 1,679 remaining respectively after LULU curation. Taxonomic annotation identified 622 OTUs from the 18S rRNA dataset against the SILVA database and 481 OTUs from the COI dataset against the MIDORI database.

Taxonomic data from WoRMS could be retrieved for 200 of the annotated COI OTUs and 190 of the 18S OTUs.

3.2. Biodiversity detection

The effect of different water eDNA sample preservation techniques differed between the target amplicons. The 18S rRNA amplicon produced significantly more OTUs in samples preserved by freezing compared to Longmire's preservation method, while the COI amplicon showed no significant difference between preservation treatments (see Appendix B for details). As a conservative approach all subsequent analyses used sample data from frozen samples. The minimum number of reads per sample was 137,624 for the COI dataset and 117,915 for the 18S dataset and so samples were rarefied to this number of reads. More OTUs in total were detected in the sediment samples compared to the water samples across all sites and both markers as shown in Figure 1b,d. In all cases, both water and sediment samples detected unique OTUs but the mean proportion of unique OTUs detected in water was lower (49.2%) in comparison to sediment (73.8%). A 2-way ANOVA testing the effect of eDNA type of environmental sample on number of OTUs generated indicated a significant effect ($p<0.001$) of sample type for both 18S rRNA and COI (See Appendix C for full model output). Ordination plots of Bray-Curtis dissimilarity (Fig. 1c,e) showed that OTUs in eDNA found in sediment and water differ in community structure as much as among sites in ordination space. Additionally, the PERMANOVA model indicated highly significant differences ($p<0.001$) between sites and eDNA medium in both the 18S rRNA and COI datasets. Furthermore, eDNA detection medium in the PERMANOVA model explained 23.2% and 32.5% of the variation in the 18S and COI data respectively, while the site explained 34.2% and 30.5% in the COI and 18S rRNA data (See Appendix D for full model output). At phylum level (Figure 2), taxonomy does not perfectly predict medium of detection, however a binomial goodness of fit test showed non-random

detection proportions in the Nematoda (Bonferroni corrected $p<0.001$), with eDNA detections mostly in sediment.

3.3. Detection of non-indigenous species

In total 22 NIS to the study region and 29 species documented as NIS in other countries were detected across the four sites (see Appendix F for full list). Out of the detected NIS seven were present in the list of 21 NIS previously detected in manual non-native invertebrate surveys at the sites. As shown in Figure 3 the results of the eDNA surveys closely matched those of the manual survey results. Four detections differed from the manual surveys, a single eDNA detection not seen in RA Surveys and three RA detections not seen in eDNA surveys (Figure 3). Remapping of cleaned reads from sites with incongruent detections to respective COI regions (Genbank Accessions: *Austrominius modestus* KY607884; *Bugula neritina* KY235450; *Ficopomatus enigmatus* KX840011) found hits for *Bugula neritina* only (5 reads from a single replicate). These reads were lost during data filtering and so did not feature in the final dataset. A detection of note was 199 reads from the sediment of Site TQ mapping to an OTU corresponding with *Arcuatula senhousia* (Asian date mussel), a novel NIS for the UK. Targeted visual surveys on tidal mudflats within two kilometres of Marina TQ confirmed the presence of this species in proximity to the sampling site. Furthermore, we generated COI sequences from these tissue samples (Genbank Accession: MH924820 and MH924821) and matched to known *A. senhousia* sequences confirming the eDNA detection of this species (see Appendix E for details of DNA barcoding). Additionally, the nematode *Cephalothrix simula* and the oligochaete *Paranais frici* were also detected using eDNA at site TQ. Both are novel species introductions to the United Kingdom, previously undocumented in academic literature.

4. Discussion

We demonstrated that eDNA from sediment and water samples reveal very different community composition, suggesting that the collection of multiple sample types provides the most comprehensive assessment of community composition. We also found that eDNA metabarcoding shows concordance with published biodiversity surveys for the detection of NIS. Furthermore, we demonstrated that eDNA metabarcoding can detect novel species introductions, suggesting that eDNA surveys are an effective tool to significantly reduce the time between introduction and detection of NIS, and in turn increase the likelihood of successful control and eradication. Our study shows how different environmental sample types can affect our understanding of both whole community composition and particular species of concern (e.g. NIS).

The majority of research using eDNA to detect aquatic macrofauna has focused on the collection of water samples, while sediment samples have received comparatively less attention. Moreover, sediment has been shown to harbour 8-1800 times more eDNA compared to water samples in freshwater ecosystems (Turner, Uy & Everhart 2015). Here we found dramatic differences in species richness in sediment and water samples, observing a consistently greater number of OTUs detected in sediment compared to water. However, our results indicated the opposite trend when considering only fish species; with more fish being detected in seawater samples compared to sediment samples (6 in water, 1 in sediment and water). Similarly, Shaw *et al.* (2016) found that sediment 12S rRNA metabarcoding detected fewer fish compared to water in a freshwater lotic environment. More broadly, taxonomy at the level of phylum did not predict if a species was detected in water, sediment or both environmental sample types (except the Nematoda, whose members are predominantly benthic inhabitants). Our study showed that at the level of phyla detection was not significantly different between sediment and water for most taxa. Similarly, we showed that for most NIS both water or sediment samples served as an excellent media for detection. Our study suggests that at lower taxonomic level the species-specific ecology of eDNA (sensu Barnes & Turner 2015) will result in convergent eDNA

occupancy in different environmental sample types, as seen in the case of the fish above. However, further work is needed to clarify how eDNA partitions into adjacent environmental samples across the tree of life.

Current eDNA metabarcoding research has identified large variation in the detected biodiversity across small spatial scales in both sediment (Nascimento *et al.* 2018) and water (O'Donnell *et al.* 2017). Additionally fractionation of environmental samples (i.e. sorting samples by particle size class) can produce significant differences in the metabarcoding results between fractions (Wangensteen *et al.* 2018a; b) indicating significant variation can be found within sites. Here we found similar patterns, with site and environmental sample type containing approximately equivalent OTU biodiversity. Future research should explore how different sample types and eDNA extraction methods affect the detection of biodiversity, especially as eDNA metabarcoding moves from an experimental technique to a routine monitoring tool (Pawlowski *et al.* 2018; Aylagas *et al.* 2018).

We found that eDNA metabarcoding of water samples accurately detects many NIS species, as seen in previous work (Borrell *et al.* 2017; Grey *et al.* 2018; Lacoursière-Roussel *et al.* 2018). In comparing our eDNA data to those collected using existing methods we found close congruence in NIS incidence. The false-negative eDNA detection of *B. neritina* was found to be a result of bioinformatic parameters, identifying that choices made during sequence processing can have an effect on the detectability of species in eDNA samples. Indeed, this has previously been shown in metabarcoding of bulk tissue samples (Scott *et al.* 2018) and work is urgently needed to determine the effects of bioinformatic parameters, variable primer binding sites and the role of DNA barcodes in reference databases for the detection of NIS in eDNA samples. The remaining incongruent detections may be as a result of community turnover among the survey dates or seasonal phenology. Indeed marine coastal communities have been shown to shift in community composition across seasons and reproductive cycles (Stachowicz & Byrnes 2006; Sutherland & Karlson 1977). It is therefore recommended to combine eDNA metabarcoding with

existing survey methods where possible, as both approaches provide reciprocal validation data. For example, important NIS may be missed in surveys based solely on eDNA (e.g. Wood *et al.* 2018), and eDNA metabarcoding may detect rare species that are often missed using other type of surveys (Blackman *et al.* 2018).

In this study we identified several NIS currently unrecorded in the United Kingdom and confirmed the eDNA detection with targeted local surveys (see Appendix E). The case of *A. senhousia* is particularly relevant as this species is spreading globally (Bachelet *et al.* 2009) and has the potential to dramatically alter benthic biodiversity when invasive (Crooks 2001; Mistri 2003). Coincidentally, field surveys along the south coast of the United Kingdom have recently confirmed the presence of *A. senhousia* (Barfield et al. 2018). The use of routine eDNA surveys has the potential to accurately assess the magnitude of the spread of *A. senhousia* along this coast, providing key information for coastal biodiversity managers. Moving forwards, the detection of NIS could be facilitated through autonomous sampling and eDNA surveys (McQuillan & Robidart 2017) to provide live species introduction data in introduction hotspots. Additionally combing these techniques with eDNA biobanking (Jarman, Berry & Bunce 2018) could provide an eDNA reference database for specific geographical regions that have high biosecurity risk, providing an invaluable resource for biodiversity managers and researchers to examine the process of invasion through time. Taken together, our study shows how effective eDNA metabarcoding is as monitoring tool of novel and resident marine species and how it allows for an unprecedented sampling replicability and accuracy of different environmental sample types.

Acknowledgements

We are grateful to John Bishop and Chris Wood from the Marine Biological Association of the United Kingdom for sharing information on marinas and their excellent NIS survey data. We thank the staff of the Environmental Sequencing Facility from the National Oceanography Centre Southampton for advice and assistance during library preparation. We acknowledge the Department of Geography & Environment from the University of Southampton for access to coring equipment and lab space. LH was supported by the Natural Environmental Research Council (grant number NE/L002531/1).

References

Ahyong, S., Costello, M.J., Galil, B.S., Gollasch, S., Hutchings, P., Katsanevakis, … Zhan, A. (2018) http://www.marinespecies.org/introduced

Airoldi, L., Turon, X., Perkol-Finkel, S., Rius, M. (2015) Corridors for aliens but not for natives: Effects of marine urban sprawl at a regional scale. *Diversity and Distributions*, **21**, 755-768.

Anderson, M.J. (2014) Permutational Multivariate Analysis of Variance (PERMANOVA). *Wiley Stats Ref: Statistics Reference Online* (eds N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri, & J.L. Teugels), pp. 1–15. John Wiley & Sons, Ltd, Chichester, UK.

Appeltans, W., Ahyong, S.T., Anderson, G., Angel, M.V., Artois, T., Bailly, N., … Costello, M.J. (2012) The magnitude of global marine species diversity. *Current Biology*, **22**, 2189–2202.

Ashton, G., Boos, K., Shucksmith, R. & Cook, E. (2006) Rapid assessment of the distribution of marine non-native species in marinas in Scotland. *Aquatic Invasions*, **1**, 209–213.

Aylagas, E., Borja, Á., Muxika, I. & Rodríguez-Ezpeleta, N. (2018) Adapting metabarcoding-based benthic biomonitoring into routine marine ecological status assessment networks. *Ecological Indicators*, **95**, 194–202.

Bachelet, G., Blanchet, H., Cottet, M., Dang, C., de Montaudouin, X., de Moura Queirós, A., Gouillieux, B. & Lavesque, N. (2009) A round-the-world tour almost completed: first records of the invasive mussel *Musculista senhousia* in the north-east Atlantic (southern Bay of Biscay). *Marine Biodiversity Records*, **2**, 2002–2005.

Barfield, P., Holmes, A., Watson, G., Rowe, G. (2018) First Evidence of *Arcuatula senhousia* (Benson, 1842), The Asian Date Mussel in UK waters. *The Journal of Conchology*, **43**, 217-222.

Barnes, M.A. & Turner, C.R. (2015) The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, **17**, 1–17.

Bax, N., Williamson, A., Aguero, M., Gonzalez, E. & Geeves, W. (2003) Marine invasive alien species: a threat to global biodiversity. *Marine Policy*, **27**, 313–323.

Bishop, J., Wood, C., Yunnie, A. & Griffiths, C. (2015) Unheralded arrivals: non-native sessile invertebrates in marinas on the English coast. *Aquatic Invasions*, **10**, 249–264.

Bista, I., Carvalho, G.R., Walsh, K., Seymour, M., Hajibabaei, M., Lallias, D., Christmas, M. & Creer, S. (2017) Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. *Nature Communications*, **8**, 14087.

Blackman, R.C., Constable, D., Hahn, C., Sheard, A.M., Durkota, J., Hänfling, B. and Lawson Handley, L., 2017. Detection of a new non-native freshwater species by DNA metabarcoding of environmental samples--first record of *Gammarus fossarum* in the UK. *Aquatic Invasions*, ***12***(2).

Borrell, Y.J., Miralles, L., Do Huu, H., Mohammed-Geba, K. & Garcia-Vazquez, E. (2017) DNA in a bottle—Rapid metabarcoding survey for early alerts of invasive species in ports. *PloS One*, **12**, e0183347.

Butchart, S.H.M., Walpole, M., Collen, B., van Strien, A., Scharlemann, J.P.W., Almond, R.E.A., … Watson, R. (2010) Global biodiversity: indicators of recent declines. *Science*, **328**, 1164–1168.

Calewaert, J.-B., Weaver, P., Gunn, V., Gorringe, P. & Novellino, A. (2016) The European Marine Data and Observation Network (EMODnet): Your Gateway to European Marine and Coastal Data. *Quantitative Monitoring of the Underwater Environment: Results of the International Marine Science and Technology Event MOQESM´14 in Brest, France* (eds B. Zerr, L. Jaulin, V. Creuze, N. Debese, I. Quidu, B. Clement, & A. Billon-Coat), pp. 31–46. Springer International Publishing, Cham.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.

Campbell, M.L., Gould, B. & Hewitt, C.L. (2007) Survey evaluations to assess marine bioinvasions. *Marine Pollution Bulletin*, **55**, 360–378.

Clarke Murray, C., Pakhomov, E.A. & Therriault, T.W. (2011) Recreational boating: a large unregulated vector transporting marine invasive species: Transport of NIS by recreational boats. *Diversity and Distributions*, **17**, 1161–1172.

Crooks, J.A. (2001) Assessing Invader Roles Within Changing Ecosystems: Historical and Experimental Perspectives on an Exotic Mussel in an Urbanized Lagoon. *Biological Invasions*, **3**, 23–36.

Dafforn, K.A., Johnston, E.L. & Glasby, T.M. (2009) Shallow moving structures promote marine invader dominance. *Biofouling*, **25**, 277–287.

Darling, J.A., Galil, B.S., Carvalho, G.R., Rius, M., Viard, F. & Piraino, S. (2017) Recommendations for developing and applying genetic tools to assess and manage biological invasions in marine ecosystems. *Marine Policy*, **85**, 56–64.

Deiner, K., Bik, H.M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., … Bernatchez, L. (2017) Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, **26**, 5872–5895.

Deiner, K., Lopez, J., Bourne, S., Holman, L.E**.,** Seymour, M., Gray, E.K., … Lodge, D.M. (2018) Optimizing the detection of marine taxonomic richness using environmental DNA metabarcoding: the effects of filter material, pore size, and extraction method. *Metabarcoding & Metagenomics, **2**, e28963*

Dougherty, M.M., Larson, E.R., Renshaw, M.A., Gantz, C.A., Egan, S.P., Erickson, D.M. & Lodge, D.M. (2016) Environmental DNA (eDNA) detects the invasive rusty crayfish *Orconectes rusticus* at low abundances. *Journal of Applied Ecology*, **53**, 722–732.

Edgar, R.C. (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, **10**, 996–998.

Fitzpatrick, M.C., Preisser, E.L., Ellison, A.M. & Elkinton, J.S. (2009) Observer bias and the detection of low-density populations. *Ecological applications: a publication of the Ecological Society of America*, **19**, 1673–1679.

Foster, V., Giesler, R.J., Wilson, A.M.W., Nall, C.R. & Cook, E.J. (2016) Identifying the physical features of marina infrastructure associated with the presence of non-native species in the UK. *Marine Biology*, **163**, 173.

Frøslev, T.G., Kjøller, R., Bruun, H.H., Ejrnæs, R., Brunbjerg, A.K., Pietroni, C. & Hansen, A.J. (2017) Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nature Communications*, **8**, 1188.

Glasby, T.M., Connell, S.D., Holloway, M.G. & Hewitt, C.L. (2007) Nonindigenous biota on artificial structures: could habitat creation facilitate biological invasions? *Marine Biology*, **151**, 887–895.

Goldberg, C.S., Turner, C.R., Deiner, K., Klymus, K.E., Thomsen, P.F., Murphy, M.A., … Taberlet, P. (2016) Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution*, **7**, 1299–1307.

Grey, E.K., Bernatchez, L., Cassey, P., Deiner, K., Deveney, M., Howland, K.L., …Lodge, D.M. (2018) Effects of sampling effort on biodiversity patterns estimated from environmental DNA metabarcoding surveys. *Scientific Reports*, **8**, 8843.

Jarman, S.N., Berry, O. & Bunce, M. (2018) The value of environmental DNA biobanking for long-term biomonitoring. *Nature Ecology & Evolution*, **2**, 1192–1193.

Kim, P., Kim, D., Yoon, T.J. & Shin, S. (2018) Early detection of marine invasive species, *Bugula neritina* (Bryozoa: Cheilostomatida), using species-specific primers and environmental DNA analysis in Korea. *Marine Environmental Research*, **139**, 1–10.

Lacoursière-Roussel, A., Howland, K., Normandeau, E., Grey, E.K., Archambault, P., Deiner, K., Lodge, D.M., Hernandez, C., Leduc, N. & Bernatchez, L. (2018) eDNA metabarcoding as a new surveillance approach for coastal Arctic biodiversity. *Ecology and Evolution*, **26**, 146.

Leray, M., Yang, J.Y., Meyer, C.P., Mills, S.C., Agudelo, N., Ranwez, V., Boehm, J.T. & Machida, R.J. (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, **10**, 34.

Lovell, S.J., Stone, S.F. & Fernandez, L. (2006) The Economic Impacts of Aquatic Invasive Species: A Review of the Literature. *Agricultural and Resource Economics Review*, **35**, 195–208.

MacConaill, L.E., Burns, R.T., Nag, A., Coleman, H.A., Slevin, M.K., Giorda, … Thorner, A.R. (2018) Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. *BMC Genomics*, **19**, 30.

Machida, R.J., Leray, M., Ho, S.-L. & Knowlton, N. (2017) Metazoan mitochondrial gene sequence reference datasets for taxonomic assignment of environmental samples. *Scientific Data*, **4**, 170027.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, **17**, 10–12.

Mazza, G., Tricarico, E., Genovesi, P. & Gherardi, F. (2014) Biological invaders are threats to human health: an overview. *Ethology Ecology & Evolution*, **26**, 112–129.

McQuillan, J.S. & Robidart, J.C. (2017) Molecular-biological sensing in aquatic environments: recent developments and emerging capabilities. *Current Opinion in Biotechnology*, **45**, 43–50.

Mistri, M. (2003) The non-indigenous mussel *Musculista senhousia* in an Adriatic lagoon: effects on benthic community over a ten year period. *Journal of the Marine Biological Association of the United Kingdom*, **83**, 1277–1278.

Molnar, J.L., Gamboa, R.L., Revenga, C. & Spalding, M.D. (2008) Assessing the global threat of invasive species to marine biodiversity. *Frontiers in Ecology and the Environment*, **6**, 485–492.

Nascimento, F.J.A., Lallias, D., Bik, H.M. & Creer, S. (2018) Sample size effects on the assessment of eukaryotic diversity and community structure in aquatic sediments using high-throughput sequencing. *Scientific Reports*, **8**, 11737.

Nunes, A.L., Katsanevakis, S., Zenetos, A. & Cardoso, A.C. (2014) Gateways to alien invasions in the European seas. *Aquatic Invasions*, **9**, 133–144.

O'Donnell, J.L., Kelly, R.P., Shelton, A.O., Samhouri, J.F., Lowell, N.C. & Williams, G.D. (2017) Spatial distribution of environmental DNA in a nearshore marine habitat. *PeerJ*, **5**, e3044.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R.B., … Wagner, H. (2011) vegan: Community ecology package. *R package version*, 117–118.

Oliver, I. & Beattie, A.J. (1993) A Possible Method for the Rapid Assessment of Biodiversity. *Conservation Biology*, **7**, 562–568.

Pawlowski, J., Kelly-Quinn, M., Altermatt, F., Apothéloz-Perret-Gentil, L., Beja, P., Boggero, A.,… Kahlert, M. (2018) The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. *The Science of the Total Environment*, **637-638**, 1295–1310.

Pearce, F., Peeler, E. & Stebbing, P. (2012) *Modelling the Risk of the Introduction and Spread of Non-Indigenous Species in the UK and Ireland; Project Report for E5405W*. CEFAS.

Pérez-Portela, R., Arranz, V., Rius, M. & Turon, M. (2013) Cryptic speciation or global spread? The case of a cosmopolitan marine invertebrate with limited dispersal capabilities. *Scientific Reports*, **3**, 3197.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. & Glöckner, F.O. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, **41**, D590–6.

R Core Team. (2018) *R: A Language and Environment for Statistical Computing*. ISBN 3-900051-07-0.

Renshaw, M.A., Olds, B.P., Jerde, C.L., Mcveigh, M.M. & Lodge, D.M. (2015) The room temperature preservation of filtered environmental DNA samples and assimilation into a phenol-chloroform-isoamyl alcohol DNA extraction. *Molecular Ecology Resources*, **15**, 168–176.

Ricciardi, A., Hoopes, M.F., Marchetti, M.P. & Lockwood, J.L. (2013) Progress toward understanding the ecological impacts of nonnative species. *Ecological Monographs*, **83**, 263–282.

Rius, M., Teske, P.R., Maríquez, P.H., Suárez-Jiménez, R., McQuaid, C.D. & Castilla, J. (2017) Ecological dominance along rocky shores, with a focus on intertidal ascidians. *Oceanography and Marine Biology*, **55**: 55-84.

Rivero, N.K., Dafforn, K.A., Coleman, M.A. & Johnston, E.L. (2013) Environmental and ecological changes associated with a marina. *Biofouling*, **29**, 803–815.

Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, **4**, e2584.

Sala, E. & Knowlton, N. (2006) Global Marine Biodiversity Trends. *Annual Review of Environment and Resources*, **31**, 93–122.

Sassoubre, L.M., Yamahara, K.M., Gardner, L.D., Block, B.A. & Boehm, A.B. (2016) Quantification of Environmental DNA (eDNA) Shedding and Decay Rates for Three Marine Fish. *Environmental Science & Technology*, **50**, 10456–10464.

Scott, R., Zhan, A., Brown, E.A., Chain, F.J.J., Cristescu, M.E., Gras, R. & MacIsaac, H.J. (2018) Optimization and performance testing of a sequence processing pipeline applied to detection of nonindigenous species. *Evolutionary Applications*, **11**, 891–905.

Sellers, G.S., Di Muri, C., Gómez, A. & Hänfling, B. (2018) Mu-DNA: a modular universal DNA extraction method adaptable for a wide range of sample types. *Metabarcoding and Metagenomics*, **2**, e24556.

Shaw, J.L.A., Clarke, L.J., Wedderburn, S.D., Barnes, T.C., Weyrich, L.S. & Cooper, A. (2016) Comparison of environmental DNA metabarcoding and conventional fish survey methods in a river system. *Biological Conservation*, **197**, 131–138.

Simpson, T.J.S., Dias, P.J., Snow, M., Muñoz, J. & Berry, T. (2017) Real-time PCR detection of *Didemnum perlucidum* (Monniot, 1983) and *Didemnum vexillum* (Kott, 2002) in an applied routine marine biosecurity context. *Molecular Ecology Resources*, **17**, 443–453.

Spens, J., Evans, A.R., Halfmaerten, D., Knudsen, S.W., Sengupta, M.E., Mak, S.S.T., Sigsgaard, E.E. & Hellström, M. (2016) Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter. *Methods in Ecology and Evolution*, **8**, 635-645.

Stachowicz JJ, Byrnes JE (2006) Species diversity, invasion success, and ecosystem functioning: disentangling the influence of resource competition, facilitation, and extrinsic factors. *Marine Ecology Progress Series,* **311**, 251-262.

Sutherland, J.P. and Karlson, R.H., 1977. Development and stability of the fouling community at Beaufort, North Carolina. *Ecological Monographs*, ***47***(4), 425-446.

Thomsen, P.F. and Willerslev, E., 2015. Environmental DNA–An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, **183**, pp.4-18.

Turner, C.R., Barnes, M.A., Xu, C.C.Y., Jones, S.E., Jerde, C.L. & Lodge, D.M. (2014) Particle size distribution and optimal capture of aqueous macrobial eDNA. *Methods in Ecology and Evolution*, **5**, 676–684.

Turner, C.R., Uy, K.L. & Everhart, R.C. (2015) Fish environmental DNA is more concentrated in aquatic sediments than surface water. *Biological Conservation*, **183**, 93–102.

Wangensteen, O.S., Cebrian, E., Palacín, C. & Turon, X. (2018a) Under the canopy: Community-wide effects of invasive algae in Marine Protected Areas revealed by metabarcoding. *Marine Pollution Bulletin*, **127**, 54–66.

Wangensteen, O.S., Palacín, C., Guardiola, M. & Turon, X. (2018b) DNA metabarcoding of littoral hard-bottom communities: high diversity and database gaps revealed by two molecular markers. *PeerJ*, **6**, e4705.

Wood, C.A., Bishop, J.D.D., Rennocks, L. & Crundwell, R. (2016) *RAS 2015 Non-Native Species Rapid Assessment Surveys in English Marinas (E Anglia & W Coast)*. Report to the Bromley Trust.

Wood, C.A., Bishop, J.D.D. & Yunnie, A.L.E. (2015a) *RAS 2014 Non-Native Species Rapid Assessment Surveys in English Marinas*. Report to the Bromley Trust.

Wood, C.A., Bishop, J.D.D. & Yunnie, A.L.E. (2015b) *Comprehensive Reassessment of NNS in Welsh Marinas*. Report to Natural Resources Wales.

Wood, S.A., Zaiko, A., Richter, I., Inglis, G.J. & Pochon, X. (2017) Development of a real-time polymerase chain reaction assay for the detection of the invasive Mediterranean fanworm, *Sabella spallanzanii*, in environmental samples. *Environmental science and pollution research international*, **24**, 17373–17382.

Wood, S.A., Pochon, X., Ming, W., von Ammon, U., Woods, C., Carter, M., Smith, M., Inglis, G.J. and Zaiko, A., in press. Considerations for incorporating real-time PCR assays into routine marine biosecurity surveillance programmes: a case study targeting the Mediterranean fanworm (*Sabella spallanzanii*) and club tunicate (*Styela clava*). *Genome*

Worm, B., Barbier, E.B., Beaumont, N., Duffy, J.E., Folke, C., Halpern, … Watson, R. (2006) Impacts of biodiversity loss on ocean ecosystem services. *Science*, **314**, 787–790.

WoRMS Editorial Board (2018) World Register of Marine Species, Available from http://www.marinespecies.org at VLIZ. Accessed 2018-11-16. doi:10.14284/170

Zhan, A., Hulák, M., Sylvester, F., Huang, X., Adebayo, A.A., Abbott, C.L., … Macisaac, H.J. (2013) High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. *Methods in Ecology and Evolution,* **4**, 558–565.

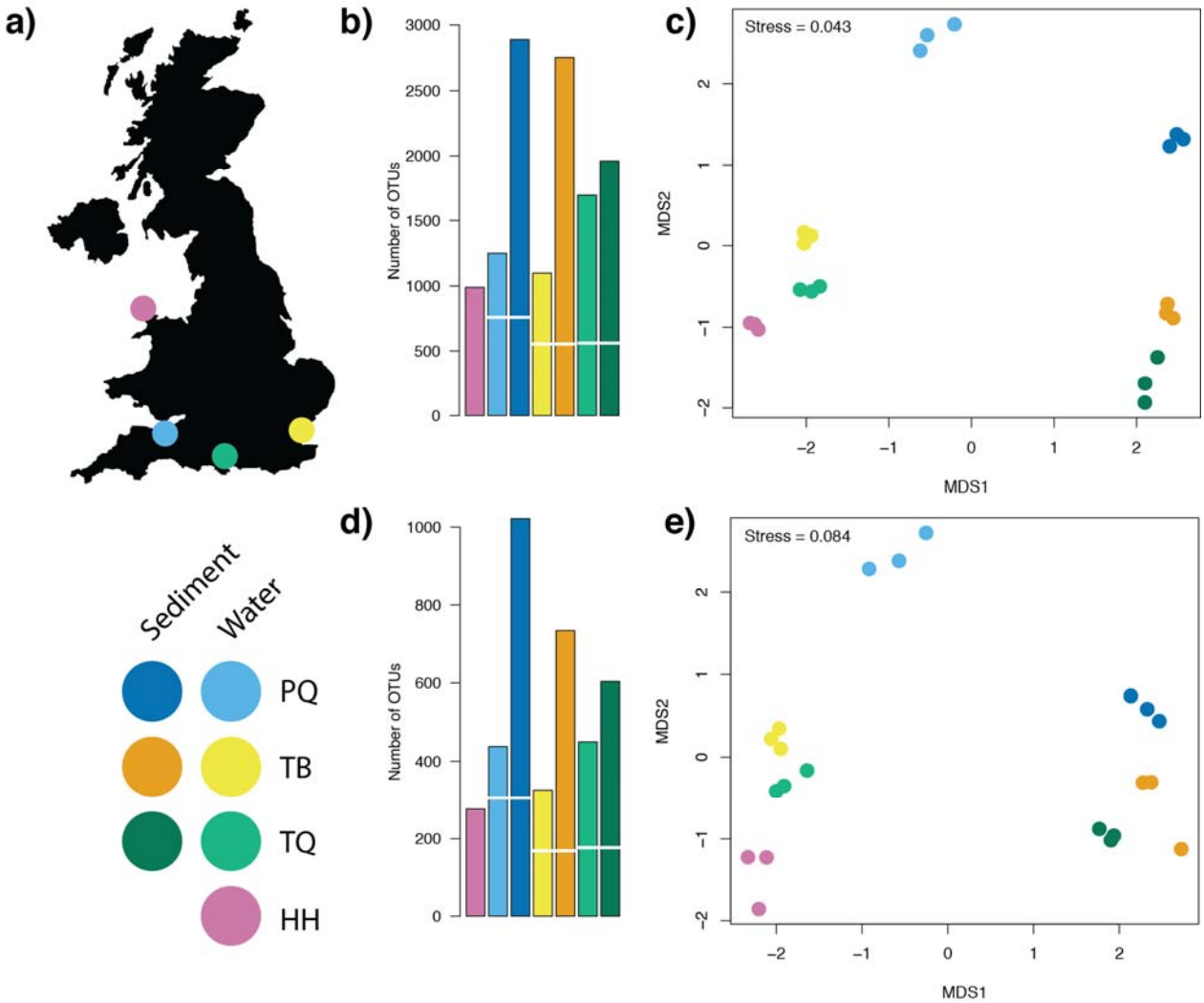

Fig. 1. *a) Map of United Kingdom indicating the geographic position of the sampled sites, a legend is provided below indicating the four sites (PQ, TB, TQ and HH) and colours for water and sediment eDNA samples for each site. Barplots detailing number of OTUs detected across sampling sites and eDNA sample type for COI b) and 18S rRNA d) metabarcoding of UK marinas, the break in bars indicates the number of shared OTUs between sediment and water eDNA samples. Non-metric multidimensional scaling ordination plots based on Bray-Curtis dissimilarities between: c) COI and e) 18S rRNA metabarcoding of marina sediment and water eDNA samples.*

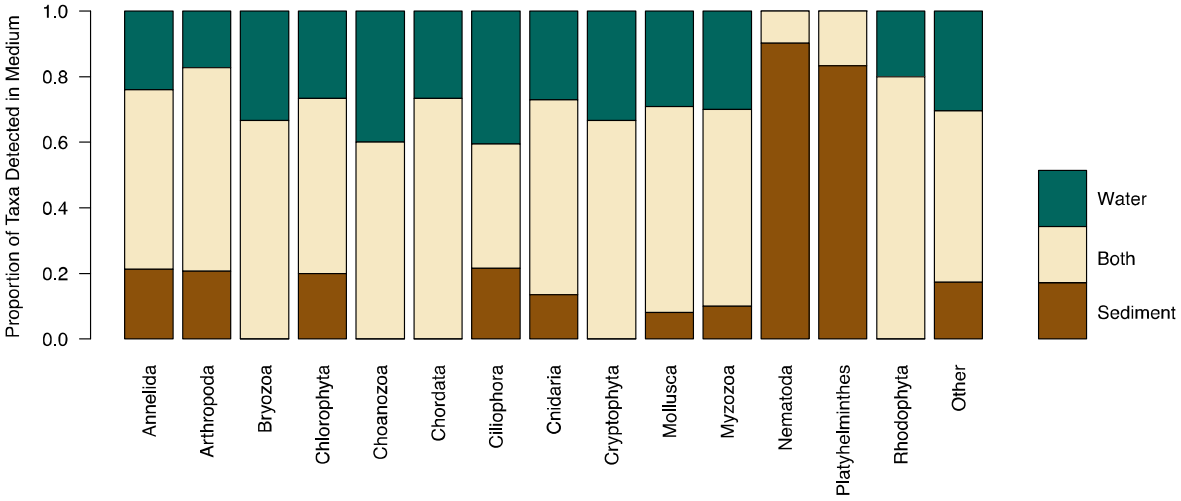

Fig. 2. *Horizontal stacked barchart detailing proportion of OTUs detected in eDNA from sediment, water or both sediment and water across the top 14 phyla for pooled 18S rRNA and COI metabarcoding of UK marinas.*

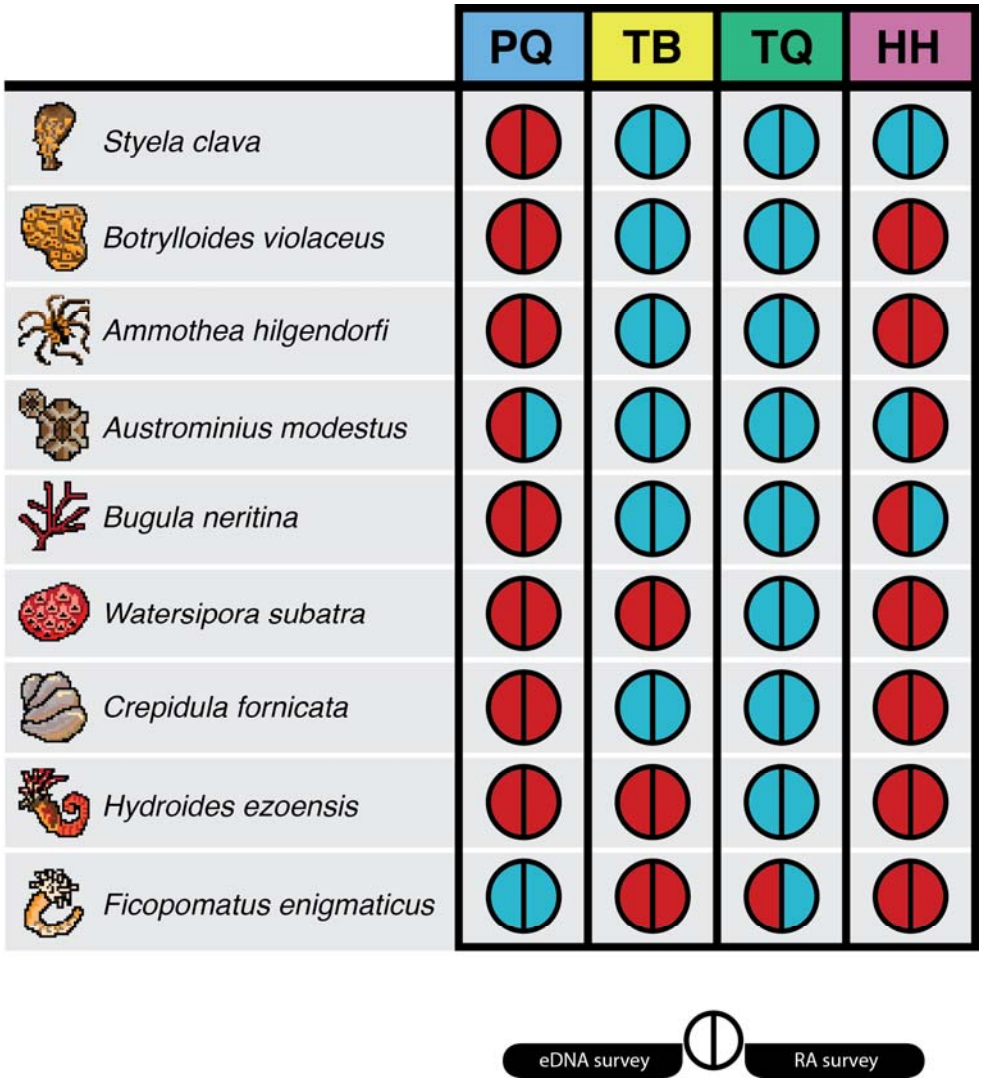

Fig. 3. *Presence absence diagram for seven non-indigenous species across four sampling sites. For each species-location the left semi-circle indicates the detection using eDNA metabarcoding surveys of 18S rRNA and COI fragments, and the right semi-circle indicates the detection from Rapid Assessment surveys. Blue indicates a positive detection for that species-location and red indicates no detection.*