The detection of novel and resident marine non-indigenous species using environmental DNA metabarcoding of seawater and sediment

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] 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

1. Biodiversity monitoring in aquatic ecosystems is challenging as it requires taxonomic expertise and is difficult to automate. One expedient approach is the use of environmental DNA (eDNA) surveys to infer species incidence from nucleic acids found in environmental samples. An advantage of eDNA surveys is that they allow rapid detection of non-indigenous species (NIS), lessening the time from introduction to discovery and increasing the likelihood of successful control and / or eradication. Despite this, relatively few studies have compared eDNA and traditional surveys for the study of NIS, or examined the differences between metabarcoding of different environmental sample types.

2. We evaluated the ability of eDNA to detect a broad range of native and NIS in urban coastal environments, and compared the results with previously published traditional surveys. We collected water and sediment samples from the same sites and then performed eDNA metabarcoding of 18S rRNA and COI genes.

3. We found very different patterns of biodiversity in water and sediment samples; with sediment containing more than two times the number of operational taxonomic units (OTUs) than seawater in some cases. The findings showed the presence of as much variation in assemblage diversity among environmental sample types as amongst geographically-segregated sampling sites. Additionally, species detection within phyla was not consistent in water or sediment samples, indicating that at a broad scale sample type does not perfectly predict taxa detected.

4. We found almost perfect agreement in species detection from eDNA and traditional surveys. Additionally, eDNA metabarcoding detected three previously undocumented species introductions. Finally, our work provided a novel high-resolution biodiversity dataset for urban marine environments.

5. *Synthesis and applications* Our study showed that the type of eDNA sample dramatically affects the detected biodiversity and that eDNA metabarcoding is accurate for the detection of

notorious NIS. Natural resource managers, ecological practitioners and researchers should consider the benefits of integrating molecular tools such as eDNA into applied ecology.

Keywords

18S, COI; eDNA; sediment; metabarcoding; non-indigenous; non-native.

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 tools that rapidly and accurately monitor species diversity. Traditional 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 limitations of traditional surveys is the use of nucleic acids found in environmental samples, such as water or sediment, to infer presence or absence of living organisms in the local environment. Such 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 detection from water samples using targeted (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 specific considerations to ensure validity, Goldberg *et al.* 2016) and can complement traditional 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 the finding that sediment has been shown to harbour 8-1800 times more eDNA compared to water samples in freshwater ecosystems (Turner, Uy & Everhart 2015) relatively few eDNA studies incorporate multiple environmental sample types (Shaw *et al.* 2016). Furthermore, no work to date has compared

species richness estimates derived from metazoan target eDNA metabarcoding between seawater and marine sediment samples.

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 are well documented and they can pose a severe threat to natural ecological systems, agriculture, biodiversity and human health (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, aquaculture, the construction of canals interconnecting large water bodies and capture aquarium industry (Molnar *et al.* 2008; Nunes *et al.* 2014). At fine (10s of km) geographical scales, other vectors such as recreational boating significantly enhance the spread and subsequent impact of NIS (Clarke Murray *et al*. 2011). Along coastal areas, NIS studies have highlighted the importance of monitoring marinas and harbours (Ashton *et al.* 2006), as these are hotspots of NIS. In these habitats, sessile marine 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.

The detection of NIS has been performed traditionally by surveys standardised by time or by reaching a species discovery asymptote (Ashton *et al.* 2006; Campbell, Gould & Hewitt 2007; Bishop *et al.* 2015). Moreover, traditional NIS surveys require expert taxonomic skills that are inherently time consuming and are not typically amenable to automation (Darling *et al.* 2017). In addition, the results of such surveys reflect only species that are being targeted at the time, with no ability to retrospectively separate erroneously grouped species in light of new discoveries.

126 This is particularly important considering that between 9,000-35,000 marine species (2.7% of
127 the total number of estimated marine species) are cryptic (i.e., morphologically similar but
128 genetically distinct) (Appeltans *et al.* 2012). Indeed, many common global sessile marine
129 invaders are morphologically indistinguishable and contain cryptic lineages as revealed by
130 genetic studies, highlighting the need for the use of genetic tools to accurately assess species
131 invasions (Pérez-Portela *et al.* 2013; Rius *et al.* 2017). The use of eDNA as monitoring tool of
132 NIS has great potential, mainly because it tackles limitations of the existing biodiversity
133 monitoring tools.
134
135 Here we tested the efficacy of eDNA metabarcoding for exploring the biodiversity of several
136 distinct marinas. We first documented the differences in both alpha and beta diversity of different
137 eDNA sample types (seawater and sediment). We then compared the eDNA metabarcoding
138 results with traditional methods and identified a number of key NIS that are both introduced in
139 the study region and / or elsewhere. Subsequently, we identified a number of previously
140 unrecorded NIS in the study region. Finally we discuss the strengths and weaknesses of eDNA
141 metabarcoding for detecting marine NIS.
142
143
144 **Methods**
145
146 *Study sites*
147 Four marinas were selected from around the United Kingdom to represent variation in modelled
148 invasion potential (Pearce, Peeler & Stebbing 2012), known NIS present (Bishop *et al.* 2015)
149 and surrounding benthic habitat type (Calewaert *et al.* 2016). Importantly, all chosen marinas
150 have been surveyed previously for NIS and so there is a good understanding of the expected
151 NIS in these areas (Wood, Bishop & Yunnie 2015a; b; Wood *et al.* 2016). Marina access was

152 contingent on anonymity and so marina names and exact locations are not presented, Fig.1a
153 shows approximate locations. Marina *TQ* is an open marina subject to tides and varying salinity
154 located in Southampton Water on the north coast of the English Channel. Marina *PQ* is a loch
155 marina open during high tide to the Bristol Channel and the Celtic Sea. Marina *TB* is located at
156 the mouth of the River Blackwater open to the North Sea. Marina *HH* is located on the Isle of
157 Anglesey and is open to the Celtic Sea.
158
159 *Environmental DNA sampling*
160 Surveys were conducted during May 2017. A total of 24 sampling points were randomly selected
161 within each site. At each point 50ml of water was collected from 10cm below the surface using a
162 sterile 60ml luer lock syringe and filtered through a 0.22mm polyethersulfone Sterivex filter
163 (Merck Millipore, Massachusetts USA). After collecting seawater from eight locations (400ml
164 total volume) the filter was changed, resulting in a total of three filters per site. To test the effect
165 of different sample preservation methods sampling was performed in duplicate. One set of three
166 filters had ~1.5ml sterile Longmire's solution (100mM Tris,10mM EDTA, 10mM NaCl, 0.5%
167 SDS) applied in the inlet valve (Renshaw et al. 2015). The second set of three filters were kept
168 on ice for no longer than eight hours before being frozen at -20°C. During the surveys, a
169 sediment sample was collected at the start and then after every 3rd water sample, for a total of 9
170 per site. We used a UWITEC Corer (UWITEC, Mondsee, Austria) to collect a sediment core
171 (600mm tall x 60mm diameter). Using a sterile disposable spatula, a subsample of 10-20g of
172 sediment was taken from the top 2cm of the core, taking care to avoid sampling the sides of the
173 core. The subsample was stored in a sterile plastic zip container and kept on ice for no longer
174 than eight hours before begin frozen at -80°C. Due to equipment malfunction no sediment
175 sample could be taken for Site HH. Disposable gloves were changed after collection of each
176 sample. All reused equipment was washed thoroughly and soaked in 10% bleach between sites,
177 before rinsing in DNAse-free sterile water.

178

179 *eDNA extraction*

180 DNA extractions were performed in a PCR–free cleanroom, separate from main laboratory
181 facilities. No high copy templates, cultures or amplicons were permitted in this sterile laboratory.
182 DNA extractions from water samples followed Spens et al. (2016) using the $SX_{CAPSULE}$ method.
183 Briefly, preservative solution was removed from the outlet and filters were dried at room
184 temperature for two hours, 720µl Qiagen buffer ATL (Qiagen, Hilden, Germany) and 80µl
185 Proteinase K was added to the filter and all samples were digested overnight at 56ºC. After
186 digestion, samples were processed using the Qiagen DNeasy Blood and Tissue Kit as per
187 manufacturer instructions. The final elution was 200µl PCR grade water.
188 Sediment extractions were performed using the Qiagen DNeasy Powermax Soil Kit using the
189 manufacturer recommended protocol. For each site the nine samples were mixed to form three
190 pooled samples; for each extraction, 10g of pooled sample was processed. A total of ten
191 samples were processed, three each per site with a single no blank, as per manufacturer's
192 instructions.

193

194 *Inhibition testing*

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

201

202 *Primer selection & library preparation*

203 Two sets of primers were chosen for metabarcoding the environmental samples: a 313bp
204 section of the standard DNA barcoding region of the cytochrome c oxidase subunit I gene (COI)
205 using primers described in Leray et al. (2013); and a variable length target of the hypervariable
206 v4 region of the nuclear small subunit ribosomal DNA (18S) using primers from Zhan *et al.*
207 (2013). Sequencing libraries were prepared using a 2-step PCR approach as detailed in Bista *et*
208 *al*. (2017), this method amplifies the target region in PCR 1 annealing universal adapters onto
209 which sample specific indices and sequencing primers are annealed in PCR 2. In contrast to
210 Bista *et al*. (2017) we used unique dual-matched indexes for PCR 2 to avoid index crosstalk
211 associated with combinatorial indexing (MacConaill et al. 2018). PCR 1 was prepared in a PCR–
212 free room, separate from main laboratory facilities. PCR conditions and reaction volumes are
213 detailed in Supplementary Information 1. Blank filters, DNA extraction kits and positive controls
214 where collected, extracted and sequenced as the experimental treatments (detailed in
215 Supplementary Information 1). Samples were pooled at an equimolar ratio and sequenced using
216 the Illumina MiSeq instrument (Illumina, San Diego, USA) with a V3 2 x 300bp kit.
217
218 *Bioinformatic analyses*
219 Samples were demultiplexed using the Illumina MiSeq control software (v 2.6.2.1). The
220 demultiplexed data was analysed using a custom pipeline written in the R programming
221 language (R Core Team 2018) hosted at https://github.com/leholman/metabarTOAD, the steps
222 are as follows. Forward and reverse paired end reads were merged using the -fastq_mergepairs
223 option of USEARCH v.10.0.240 (Edgar 2013) with maximum difference of 15, percent identity of
224 80% and quality filter set at maximum expected errors of 1. Both the forward and reverse primer
225 sequences were matched using Cutadapt v.1.16 (Martin 2011) and only sequences containing
226 both primer regions were retained. Sequences were discarded if they were outside of a defined
227 length boundary (303-323bp for COI, 375-450bp for 18S) using Cutadapt. Sequences were then
228 pooled, singletons were discarded and sequences were quality filtered with a maximum

229 expected error of 1 using the -fastq_filter option of vsearch v.2.4.3 (Rognes *et al.* 2016).
230 Sequences were then denoised and chimeras filtered using the unoise3 algorithm implemented
231 in USEARCH. The resultant OTUs were curated using the LULU package v.0.1.0 in R (Frøslev
232 *et al.* 2017). An OTU by sample table was produced by mapping the merged and trimmed reads
233 against the curated OTUs, reporting a single best hit within 97% of the query sequence. The
234 OTU x sample table was filtered in R as follows. To minimise the chance of spurious OTUs
235 being included in the final dataset any record with less than 3 raw reads were changed to zero
236 and any OTU that did not appear in more than one sample was removed from the analysis.
237 OTUs found in negative controls were removed from the analysis.
238
239 *Taxonomic assignment*
240 Assigning correct taxonomy to an unknown set of marine sequences can be challenging as
241 large databases require vast computational resources for query matching; many databases
242 contain errors and the taxonomy of some marine groups is uncertain. With such limitations in
243 mind, we assigned taxonomy using a BLAST v.2.6.0+ search (Camacho *et al.* 2009) returning
244 the single best hit from databases within 97% of the query. The MIDORI database
245 (UNIQUE_20180221) (Machida *et al.* 2017) was used for the COI data and the SILVA database
246 (SSU r132) (Quast *et al.* 2013) was used for the 18S rRNA data. The match taxa tool from the
247 World Register of Marine Species (WoRMS) (WoRMS Editorial Board 2018) was used to filter
248 the data for marine species and check the classification. Remaining annotations were checked
249 against the World Register of Introduced Marine Species (WRIMS) (Ahyong *et al.* 2018) to
250 determine non-indigenous status.
251
252 *Statistical analyses*
253 All statistical analyses were conducted in R v3.5.0. The Vegan R package v.2.5.2 (Oksanen *et*
254 *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 linear regression model following the formula *number_of_OTUs~sedimentorwater*site* implemented in R using the function lm(). 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 *dissimlairty_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 up 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.

Results

Raw sequencing results & OTU generation

280 Sequencing produced a total of 17.8 million paired end reads, with 15.2 million sequences
281 remaining after paired end read merging and quality filtering. The average number of sequences
282 per sample after filtering (excluding control samples) was 200,185 ± 64,019 (s.d). No template
283 control samples contained an average of 811 ± 3,402 (s.d) sequences. One control sample
284 contained ~15,000 sequences that mapped to an operational taxonomic unit (OTU) that had
285 100% identity match to a sequence of a terrestrial fungi (Genbank: FJ804151.1), excluding this
286 OTU gives an average of 51 ± 94 (s.d) sequences per no-template control sample. Denoising
287 produced 8,069 OTUs for COI and 2,433 for 18S with 6,435 and 1,679 remaining respectively
288 after LULU curation. Taxonomic annotation identified 395 OTUs from the 18S rRNA dataset
289 against the SILVA database and 219 OTUs from the COI dataset against the MIDORI database.
290 Taxonomic data from WoRMS could be retrieved for 204 of the annotated COI OTUs and 138 of
291 the 18S OTUs.
292
293 *Biodiversity detection*
294
295 The effect of different water eDNA sample preservation techniques differed between the target
296 amplicons. The 18S rRNA amplicon produced significantly more OTUs in samples preserved by
297 freezing compared to Longmire's preservation method, while the COI amplicon showed no
298 significant difference between preservation treatments (see Supplementary Information 2 for
299 details). As a conservative approach all subsequent analyses used sample data from frozen
300 samples. The minimum number of reads per sample was 137,624 for the COI dataset and
301 117,915 for the 18S dataset and so samples were rarefied to this number of reads. More OTUs
302 in total were detected in the sediment samples compared to the water samples across all sites
303 and both markers as shown in Figure 1b,d. In all cases, both water and sediment samples
304 detected unique OTUs but the mean proportion of unique OTUs detected in water was lower
305 (49.2%) in comparison to sediment (73.8%). A 2-way ANOVA testing the effect of eDNA type of

306 environmental sample on number of OTUs generated indicated a significant effect ($p<0.001$) of
307 sample type for both 18S rRNA and COI (See Supplementary Information 3 for full model
308 output). Ordination plots of Bray-Curtis dissimilarity (Fig. 1c,e) showed that OTUs in eDNA found
309 in sediment and water differ in community structure as much as among sites in ordination space.
310 Additionally, the PERMANOVA model indicated highly significant differences ($p<0.001$) between
311 sites and eDNA medium in both the 18S rRNA and COI datasets. Furthermore, eDNA detection
312 medium in the PERMANOVA model explained 23.2% and 32.5% of the variation in the 18S and
313 COI data respectively, while the site explained 34.2% and 30.5% in the COI and 18S rRNA data
314 (See Supplementary Information 4 for full model output). At phylum level (Figure 2), taxonomy
315 does not perfectly predict medium of detection, however a binomial goodness of fit test shows
316 non-random detection proportions in the Nematoda (Bonferroni corrected $p=0.005$), with eDNA
317 detections mostly in sediment.
318
319 *Detection of non-indigenous species*
320 In total 16 NIS to the study region and 30 species documented as NIS in other countries were
321 detected across the four sites (see Supplementary Table 1 for full list). Out of the detected NIS
322 seven were present in the list of 21 NIS previously detected in manual non-native invertebrate
323 surveys at the sites. As shown in Fig. 3 the results of the eDNA surveys closely matched those
324 of the manual survey results. Only a single detection differed from the manual surveys, an eDNA
325 false-negative detection of *Bugula neritina* at Site HH. Remapping of the cleaned reads to the
326 *Bugula neritina* COI region (Genbank Accession: KY235450.1) indicated that 5 reads from a
327 single replicate corresponded with *Bugula neritina.* These reads were lost during data filtering
328 and so did not feature in the final dataset. A detection of note was 199 reads from the sediment
329 of Site TQ mapping to an OTU corresponding with *Arcuatula senhousia* (Asian Date Mussel), a
330 previously undocumented NIS for the UK. Targeted visual surveys on tidal mudflats within two
331 kilometres of Marina TQ confirmed the presence of this species in proximity to the sampling site.

332 Furthermore, COI sequences generated from these physical samples (Genbank Accession:
333 MH924820 and MH924821) matched to known *A. senhousia* sequences confirming the eDNA
334 detection (see Supplementary Information 5 for details of DNA barcoding). Additionally, the
335 nematode *Cephalothrix simula* and the oligochaete *Paranais frici* were also detected using
336 eDNA at site TQ. Both are new species introductions to the United Kingdom, previously
337 undocumented in academic literature.
338
339 **Discussion**
340
341 We demonstrated that very different community composition can be detected in seawater and
342 sediment samples, affirming the importance of integrating multiple sample types for obtaining a
343 full assessment of community composition. We also found that eDNA metabarcoding shows
344 excellent concordance with traditional methods for the detection of NIS. Finally, we
345 demonstrated that eDNA metabarcoding can detect novel species introductions. This study
346 emphasises the need to have a thorough understanding of the effects of environmental sample
347 types both at the level of whole community and for specific species of concern. eDNA
348 metabarcoding continues to be used in both regional and global biodiversity surveys.
349
350 The majority of research using eDNA to detect aquatic macrofauna is based on the collection of
351 water samples, while sediment samples have received comparatively less attention. We found
352 dramatic differences in species richness in sediment and water samples, observing a
353 consistently greater number of OTUs detected in sediment compared to water. Shaw *et al.*
354 (2016) found that sediment 12S rRNA metabarcoding detected fewer fish compared to water in
355 a freshwater lotic environment. Our results indicated a similar trend when considering only fish
356 species; with more fish being detected in seawater samples compared to sediment samples (5
357 in water, 2 in sediment and water), but the opposite when considering all OTUs. More broadly,

358 taxonomy at the level of phylum did not predict if a species will be detected in water, sediment or
359 both sample types (except the Nematoda, whose members are predominantly benthic
360 inhabitants). It is likely, as seen in the case of fish above, that at lower taxonomic levels the
361 species-specific ecology of eDNA (sensu Barnes & Turner 2015) will result in convergent eDNA
362 occupancy in different sample types. However, further work is needed to clarify how eDNA
363 partitions into adjacent environmental samples across the tree of life. Our study showed that at
364 the level of phyla detection was not significantly different between sediment and water for most
365 taxa. Similarly, we showed that for most NIS both water or sediment samples served as an
366 excellent media for detection.
367
368 Current eDNA metabarcoding research has identified large variation in the detected diversity
369 across small spatial scales in both sediment (Nascimento *et al.* 2018) and seawater (O'Donnell
370 *et al.* 2017). Additionally fractionation of environmental samples (i.e. sorting samples by particle
371 size class) can produce significant differences in the metabarcoding results between fractions
372 (Wangensteen *et al.* 2018a; b) indicating that reliable variation can be discovered at a scale of
373 site. Here we found similar patterns, with site and eDNA sample type containing approximately
374 equivalent OTU biodiversity. Future research should explore how eDNA extracted using different
375 sample types and extraction methods affects the detection of biodiversity, especially as eDNA
376 metabarcoding moves from an experimental technique to a routine monitoring tool (Pawlowski *et*
377 *al.* 2018; Aylagas *et al.* 2018).
378
379 We found that eDNA metabarcoding of water samples accurately detects many NIS species, as
380 seen in previous work (Borrell *et al.* 2017; Grey *et al.* 2018; Lacoursière-Roussel *et al.* 2018).
381 Additionally, we identified the presence of NIS in sediment samples for the first time. In
382 comparing these data to those collected using traditional methods we found almost perfect
383 agreement in NIS presence. The single false-negative was found to be a result of bioinformatic

384 parameters, identifying that choices made during sequence processing can have an effect on
385 the detectability of species in eDNA samples. Indeed, this has previously been shown in
386 metabarcoding of bulk samples (Scott *et al.* 2018) and work is urgently needed to determine
387 optimal bioinformatic parameters, the effects of primer binding sites and the role of DNA
388 barcodes in reference databases for the detection of NIS in eDNA samples. It is therefore
389 important to combine eDNA metabarcoding with traditional surveys where possible, as both
390 methods provide reciprocal validation data (e.g. important NIS may be missed using molecular
391 techniques or eDNA metabarcoding detects rare species that often missed by traditional
392 surveys). We identified several NIS currently unrecorded in the UK and confirmed the detection
393 with targeted local surveys to confirm the presence of *A. senhousia*. The case of *A. senhousia* is
394 particularly relevant when evaluating eDNA metabarcoding for NIS detection as the species is
395 spreading globally (Bachelet *et al.* 2009), has the potential to dramatically alter benthic
396 biodiversity (Crooks 2001; Mistri 2003) and may have long remained undetected. Future
397 research should evaluate the sensitivity of eDNA metabarcoding for the detection of novel
398 species introductions, which may allow an unprecedented level of accuracy during early stages
399 of invasion. The incorporation of autonomous sampling (McQuillan & Robidart 2017) and eDNA
400 biobanking (Jarman, Berry & Bunce 2018) could also help the rapid detection of NIS by
401 improving DNA reference databases for specific geographical regions that have high biosecurity
402 risk, providing an invaluable resource for biodiversity managers and researchers alike.
403
404
405
406
407
408
409

Author Contributions

L.E.H. and M.R. designed the experiment, L.E.H. collected samples, generated and analysed the data, designed all figures and wrote the first draft of the paper. All authors contributed critically to the drafts and gave final approval for publication.

Acknowledgements

We are grateful to the Bishop Lab (Marine Biological Association) and affiliates for excellent NIS survey data and for sharing information on marinas. We acknowledge staff in the Environmental Sequencing Facility at 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).

Data accessibility

Raw Illumina sequencing data is available from the European Nucleotide Archive under accession number (pending).

Associated metadata, script and intermediate files can be found on GitHub with the following DOI: 10.5281/zenodo.1453959.

References

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

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.

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

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

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

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

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

468 Butchart, S.H.M., Walpole, M., Collen, B., van Strien, A., Scharlemann, J.P.W., Almond,
469 R.E.A., … Watson, R. (2010) Global biodiversity: indicators of recent declines. *Science*,
470 **328**, 1164–1168.

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

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

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

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

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

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

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

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

495 Deiner, K., Lopez, J., Bourne, S., Holman, L.E**.,** Seymour, M., Gray, E.K., … Lodge, D.M. *In*
496 *press*. Optimizing the detection of marine taxonomic richness using environmental DNA
497 metabarcoding: the effects of filter material, pore size, and extraction method.
498 *Metabarcoding & Metagenomics*

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

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

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

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

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

513 Glasby, T.M., Connell, S.D., Holloway, M.G. & Hewitt, C.L. (2007) Nonindigenous biota on
514 artificial structures: could habitat creation facilitate biological invasions? *Marine Biology*,
515 **151**, 887–895.

516 Goldberg, C.S., Turner, C.R., Deiner, K., Klymus, K.E., Thomsen, P.F., Murphy, M.A., …
517 Taberlet, P. (2016) Critical considerations for the application of environmental DNA
518 methods to detect aquatic species ed M. Gilbert. *Methods in Ecology and Evolution*, **7**,
519 1299–1307.

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

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

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

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

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

536 Lovell, S.J., Stone, S.F. & Fernandez, L. (2006) The Economic Impacts of Aquatic Invasive
537 Species: A Review of the Literature. *Agricultural and Resource Economics Review*, **35**,
538 195–208.

539 MacConaill, L.E., Burns, R.T., Nag, A., Coleman, H.A., Slevin, M.K., Giorda, … Thorner, A.R.
540 (2018) Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index
541 cross-talk and significantly improve sensitivity of massively parallel sequencing. *BMC*
542 *Genomics*, **19**, 30.

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

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

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

550 McQuillan, J.S. & Robidart, J.C. (2017) Molecular-biological sensing in aquatic environments:
551 recent developments and emerging capabilities. *Current Opinion in Biotechnology*, **45**,
552 43–50.

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

556 Molnar, J.L., Gamboa, R.L., Revenga, C. & Spalding, M.D. (2008) Assessing the global threat
557 of invasive species to marine biodiversity. *Frontiers in Ecology and the Environment*, **6**,
558 485–492.

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

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

564 O'Donnell, J.L., Kelly, R.P., Shelton, A.O., Samhouri, J.F., Lowell, N.C. & Williams, G.D.
565 (2017) Spatial distribution of environmental DNA in a nearshore marine habitat. *PeerJ*,
566 **5**, e3044.

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

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

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

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

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

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

584 R Core Team. (2018) *R: A Language and Environment for Statistical Computing*. ISBN 3-
585 900051-07-0.

586 Renshaw, M.A., Olds, B.P., Jerde, C.L., Mcveigh, M.M. & Lodge, D.M. (2015) The room
587 temperature preservation of filtered environmental DNA samples and assimilation into a
588 phenol-chloroform-isoamyl alcohol DNA extraction. *Molecular Ecology Resources*, **15**,
589 168–176.

590 Ricciardi, A., Hoopes, M.F., Marchetti, M.P. & Lockwood, J.L. (2013) Progress toward
591 understanding the ecological impacts of nonnative species. *Ecological Monographs*, **83**,
592 263–282.

593 Rius, M., Teske, P.R., Maríquez, P.H., Suárez-Jiménez, R., McQuaid, C.D. & Castilla, J.
594 (2017) Ecological dominance along rocky shores, with a focus on intertidal ascidians.
595 *Oceanography and Marine Biology*. pp. 55–86.

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

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

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

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

605 Scott, R., Zhan, A., Brown, E.A., Chain, F.J.J., Cristescu, M.E., Gras, R. & MacIsaac, H.J.
606 (2018) Optimization and performance testing of a sequence processing pipeline applied
607 to detection of nonindigenous species. *Evolutionary Applications*, **11**, 891–905.

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

611 Shaw, J.L.A., Clarke, L.J., Wedderburn, S.D., Barnes, T.C., Weyrich, L.S. & Cooper, A.
612 (2016) Comparison of environmental DNA metabarcoding and conventional fish survey
613 methods in a river system. *Biological Conservation*, **197**, 131–138.

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

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

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

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

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

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

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

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

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

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

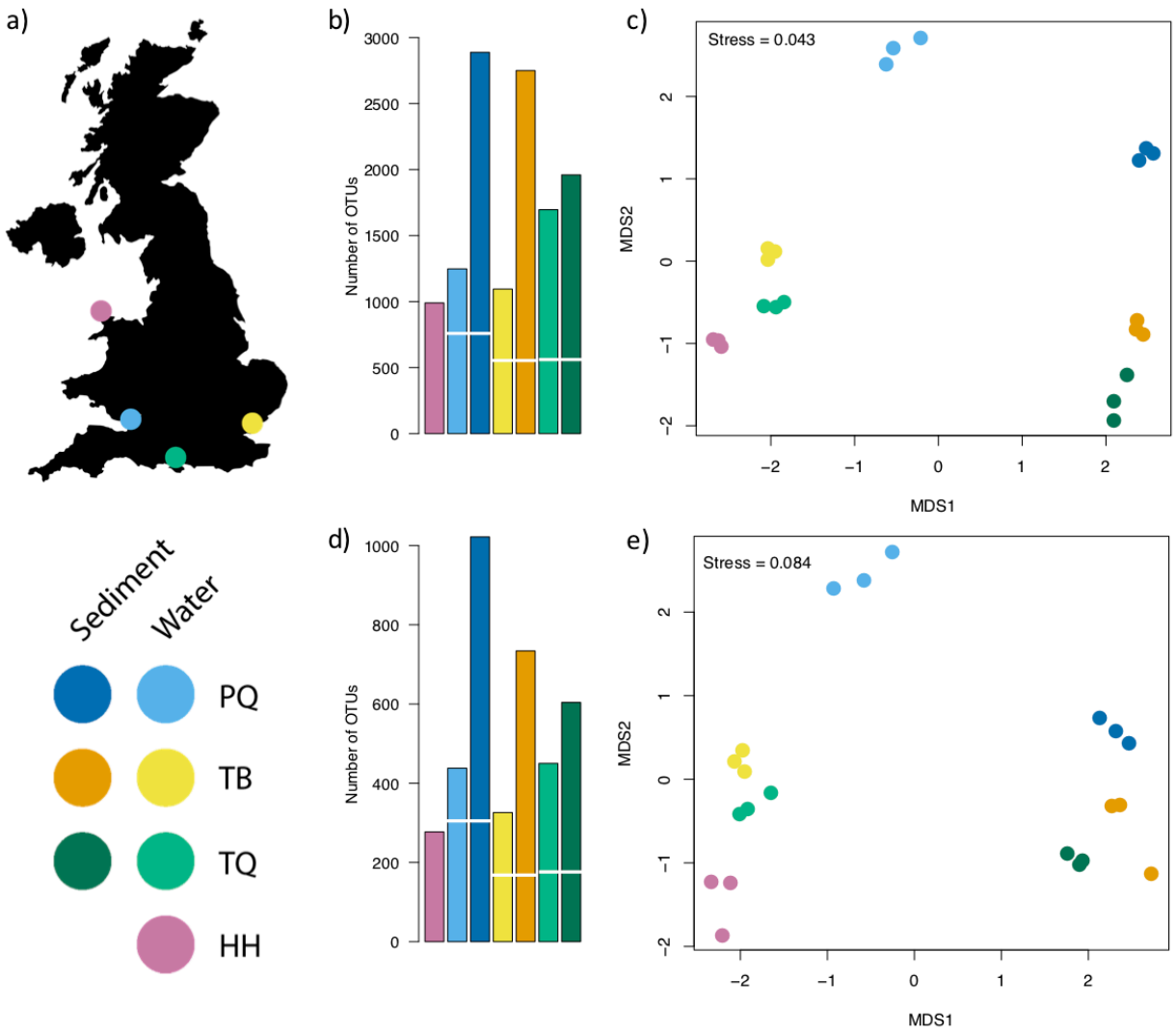
643 Worm, B., Barbier, E.B., Beaumont, N., Duffy, J.E., Folke, C., Halpern, … Watson, R. (2006)
644 Impacts of biodiversity loss on ocean ecosystem services. *Science*, **314**, 787–790.

645 WoRMS Editorial Board. (2018) World Register of Marine Species,
646 http://www.marinespecies.org

647 Zhan, A., Hulák, M., Sylvester, F., Huang, X., Adebayo, A.A., Abbott, C.L., … Macisaac, H.J.
648 (2013) High sensitivity of 454 pyrosequencing for detection of rare species in aquatic
649 communities. *Methods in Ecology and Evolution,* **4**, 558–565.

650

651



652

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

660

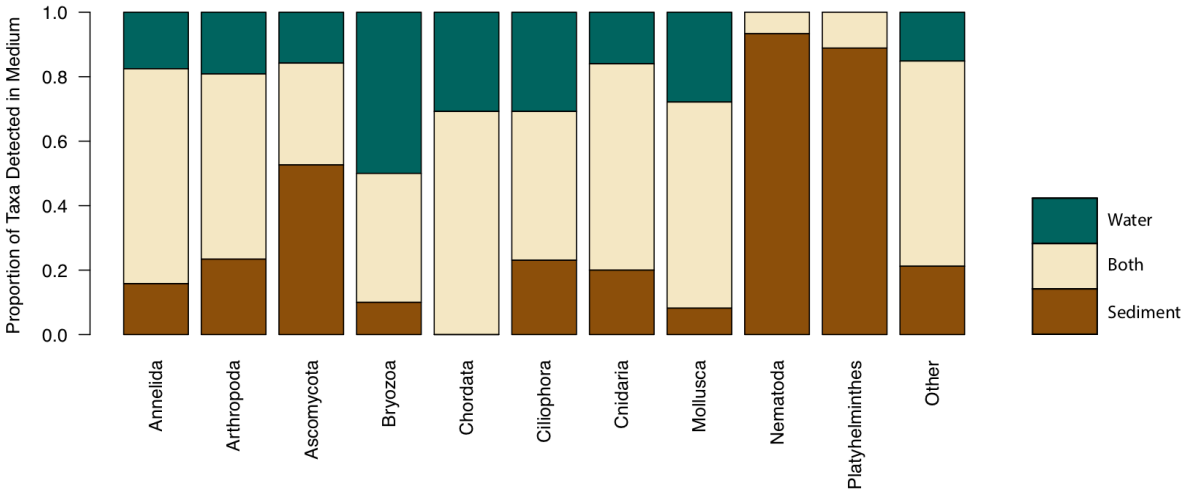


Fig. 2. *Horizontal stacked barchart detailing proportion of OTUs detected in eDNA from sediment, water or both across the top 10 phyla for pooled 18S rRNA and COI metabarcoding of UK marinas.*

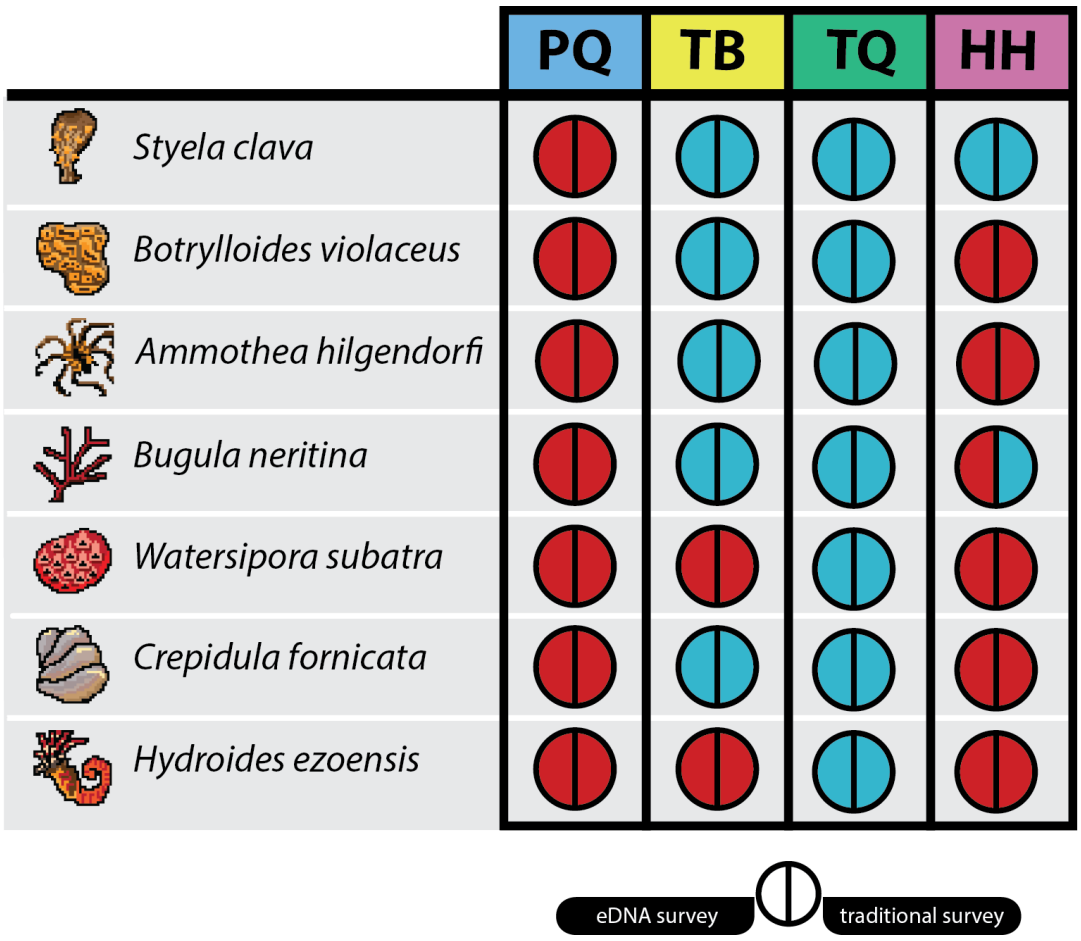


Fig. 3. *Presence absence diagram for seven non-indigenous species across four sampling sites. For each species-location the left semi-circle indicates the detection using eDNA metabarcoding surveys of 18S rRNA and COI fragments, and the right semi-circle indicates the detection from traditional manual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