Guidelines and considerations for metabarcoding-based port baseline biodiversity surveys: towards improved marine non-indigenous species monitoring

Anaïs Rey, Oihane C. Basurko, Naiara Rodriguez-Ezpeleta*

AZTI, Marine Research Division, Txatxarramendi ugartea z/g, Sukarrieta 48395 Bizkaia, Spain

Corresponding author: nrodriguez@azti.es

Running head: Metabarcoding-based port monitoring

Summary

1. Monitoring introduction and spread of non-indigenous species via maritime transport requires performing port biological baseline surveys. Yet, the comprehensiveness of these surveys is often compromised by the large number of habitats present in a port, the seasonal variability of the inhabiting communities and the time-consuming morphological approach used for taxonomic classification. Metabarcoding represents a promising alternative for rapid comprehensive port biological baseline surveys, but, before this technique can be routinely applied in this context, standardized protocols should be developed.
2. We applied metabarcoding using two alternative barcodes (based on the Cytochrome Oxidase I or the 18S ribosomal RNA gene) to about two hundred port samples collected i) from diverse habitats (water column – including environmental DNA and zooplankton, sediment and fouling structures), ii) at different sites within the port (from inner to outer estuary), and iii) during the four seasons of the year. Comparing the biodiversity metrics and taxonomic composition derived from each sample group, we investigated the impact of the different strategies for metabarcoding-based port biodiversity baseline surveys.
3. Each sampling method resulted in a distinct community profile and water samples alone using universal primers did not produce comprehensive macroorganismal biodiversity to substitute organismal sampling. Sampling at different seasons and locations resulted in higher observed biodiversity, but nestedness analyses suggested that sampling could be reduced to two seasons.

Metabarcoding also allowed to detect previously recorded non-indigenous species as well as to reveal presence of new ones, even if in low abundance.

4. We provide the first comprehensive evaluation of metabarcoding for port biodiversity baseline surveys from which we have derived guidelines for how, when and where samples should be collected. These guidelines are key to develop the cost-effective and standardizable approach needed for monitoring non-indigenous species introduction to ports via ballast water and hull fouling, particularly relevant with the recent entry into force of the International Convention for the Control and Management of Ships' Ballast Water and Sediments.

Key-words: ballast water, environmental DNA, environmental monitoring, invasive species, metabarcoding, non-indigenous species, port biological baseline surveys

48 **Introduction**

49 Globalization has led to an increased maritime transportation, with more and
50 larger ships than ever before transferring species into ports via ballast water and hull
51 fouling (Seebens, Gastner, & Blasius, 2013). Among the thousands species arriving daily
52 (Carlton, 1999), some are Non-Indigenous Species (NIS) and can become invasive,
53 unbalancing the equilibrium of the recipient ecosystem (Bax, Williamson, Aguero,
54 Gonzalez, & Geeves, 2003). Thus, being important gateways for introduction of NIS,
55 ports need to be monitored to provide information required by legal frameworks aiming
56 at controlling biological invasions (Lehtiniemi et al., 2015).

57 One of these legal frameworks is the International Convention for the Control
58 and Management of Ships' Ballast Water and Sediments (IMO, 2004), currently in force.
59 This convention requires that ships treat their ballast water before its release to port,
60 unless they show that the risk of transferring NIS between the donor and recipient ports
61 is limited (David, Gollasch, & Pavliha, 2013). Such risk assessment relies on the
62 combination of port environmental (temperature and salinity) and biological
63 information. The obtention of biological information is done by cataloguing biodiversity
64 through Port Biological Baseline Surveys (PBBS). Given the diverse range of habitats
65 (such as soft sediment, water column, or exposed and sheltered artificial structures)
66 allowing presence of distinct organismal groups (such as benthic macrofauna, fouling
67 organisms or planktonic organisms), comprehensive PBBS require collecting samples
68 using various methods (Kraus et al., 2018) and at different spatial and temporal scales
69 (Lehtiniemi et al., 2015).

70 Global initiatives have been initiated to standardize sampling for PBBS
71 (HELCOM/OSPAR, 2013; Awad, Haag, Anil, & Abdulla, 2014; Kraus et al., 2018), and are

72 based on morphological taxonomic identification of the species found. Yet, this method
73 lacks accuracy for identification of larvae and eggs, developmental stages at which many
74 NIS are transported (Gittenberger, Rensing, Niemantsverdriet, Schrieken, & Stegenga,
75 2014), and relies on taxonomists who are often specialized on the local biota, but who
76 have limited knowledge of alien taxa (Pyšek et al., 2013) and/or on specific taxonomic
77 groups coexisting in a port (Bishop & Hutchings, 2011). Moreover morphological
78 identification is time-consuming (Mandelik, Roll, & Fleischer, 2010), especially when
79 considering the several tens of samples required to characterize a port. These limitations
80 translate into PBBS being completed several years after sample collection (Bott, 2015),
81 reducing the effectiveness of prevention of NIS introduction and spread control
82 strategies.

83 DNA metabarcoding (Taberlet, Coissac, Pompanon, Brochmann, & Willerslev,
84 2012), the simultaneous identification of taxa present in a complex environmental
85 sample based on a conserved DNA fragment, is revolutionizing traditional biodiversity
86 monitoring (Creer et al., 2016; Elbrecht, Vamos, Meissner, Aroviita, & Leese, 2017), and
87 could represent the cost-effective and possible to standardize alternative required for
88 PBBS (Lehtiniemi et al., 2015). Metabarcoding can be applied to DNA extracted from
89 bulk organismal samples (*e.g.* zooplankton) or to environmental DNA (eDNA), defined
90 as traces of DNA released in the environment in form of cells, feces, skin, saliva, mucus,
91 etc. (Shaw, Weyrich, & Cooper, 2016). The later has received particular attention lately
92 for its potential to cost-effectively and noninvasively survey species richness from many
93 ecosystems (Deiner, Fronhofer, Mächler, Walser, & Altermatt, 2016), and for being able
94 to detect spatially discrete communities (Jeunen et al., 2019).

95 Metabarcoding has been described as a promising tool for NIS monitoring
96 (Comtet, Sandionigi, Viard, & Casiraghi, 2015), and recent studies have highlighted the
97 potential of metabarcoding applied to eDNA to perform port biodiversity surveys
98 (Borrell, Miralles, Huu, Mohammed-Geba, & Garcia-Vazquez, 2017; Grey et al., 2018;
99 Lacoursière-Roussel et al., 2018). Yet, very few studies have performed port surveys
100 based on additional sampling substrates (Brown, Chain, Zhan, MacIsaac, & Cristescu,
101 2016; Zaiko et al., 2016) and only one has compared biodiversity assessments obtained
102 by the different substrates (Koziol et al., 2019). Additionally, few studies have
103 considered the spatial and temporal variability of ports (*e.g.* eDNA in arctic ports
104 Lacoursière-Roussel et al. (2018) and settlement plates in austral temperate port Zaiko
105 et al.(2016)) but none have fully evaluated their impact on all the different communities
106 of a port, which is critical to ensure maximum biodiversity recovery in port baseline
107 surveys. In sum, the design of sampling schemes for metabarcoding based PBBS that
108 best encompass the heterogeneity of ports concerning substrate, locality and season
109 has not been fully investigated to date.

110 Here, we have applied metabarcoding to about two hundred samples collected
111 from the port of Bilbao (Northern Spain) to evaluate the effect of i) using various
112 sampling methods to capture biodiversity found in different substrates, ii) sampling at
113 different sites from outer to inner estuary, iii) sampling during different seasons of the
114 year, and iv) using alternative genetic markers. Our results have allowed to conclude
115 that combining the alternative sampling substrates and universal primers allows a more
116 comprehensive port biodiversity assessment, that sampling from a single substrate type
117 or at a single season does not capture the whole port biodiversity and that the port
118 salinity gradient needs to be considered when looking for NIS. From our results, we have

119 derived a series of guidelines and considerations for PBBS that will be particularly
120 relevant with recent entry into force of the International Convention for the Control and
121 Management of Ships' Ballast Water and Sediments.
122
123 **Materials and methods**
124 *Sampling*
125 Zooplankton, sediment, fouling and water samples were collected from the port of
126 Bilbao in autumn 2016, winter 2017, spring 2017, and late summer 2017 at four sites
127 (Figs 1 and S1). Sites 1, 2 and 3 depict the outer parts of the port with busy berths and
128 are characterized by deep waters (10-30 m) and salinity (>30 PSU); site 4 depicts the
129 inner, less busy part of the port and is characterized by low water depth (6-9 m) and
130 lower salinity (>20 PSU). Zooplankton, sediment and fouling sampling was based on the
131 protocol designed by Helcom/Ospar (HELCOM/OSPAR, 2013). Zooplankton samples
132 were collected in vertical tows using Pairovet nets (mesh sizes of 60 and 280 µm) at
133 three points per site, mixing the collected material to have one sample per mesh size
134 per site. The collected zooplankton was grinded with a mortar until no integer organism
135 could be appreciated and was stored in 96% ethanol at -20 °C. Sediment samples were
136 collected by sieving material collected from a Van Veen grab (0.07–0.1 m^2) with a 1 mm
137 mesh size sieve. Retained material was processed following Aylagas et al. (2016), and
138 retrieved benthic macroinvertebrates were stored in 96% ethanol at -20 °C. Fouling
139 samples were collected by placing 15 x 15 cm polyvinyl chloride plates at 1, 3 and 7 m
140 depth in each site. Plates were deployed in winter and spring and recovered in spring
141 and late summer respectively (see Table S1). At recovery, plates were placed in
142 individual sterile plastic bags, soaked in 96% ethanol and thoroughly scrapped with a

143 scalpel to retrieve fouling organisms attached to it. The detached organisms were then
144 homogenized with a blender and stored in 96% ethanol at -20 °C. At each sampling site,
145 1 L surface water sample was collected from each sampling point with a bottle. Samples
146 from the same site were combined into a single sample by site. In sites 1 and 3, three
147 additional samples were taken at maximum depth using a Niskin bottle and combined.
148 Each 3 L combined water sample was filtered using a 0.45 µm Sterivex filter unit (Merck
149 Chemicals & Life Science), which was stored at -80 °C until further use. Additionally,
150 surface water samples from three sampling sites from the ports of Vigo and A Coruña,
151 located in Galicia, Northwestern Spain, were collected in March 2017 and filtered as
152 described above.
153
154 *DNA extraction, library preparation and sequencing*
155 Total genomic DNA was isolated from zooplankton samples with the DNeasy blood &
156 tissue kit (QIAGEN), benthic macroinvertebrate and fouling organism samples with,
157 respectively, PowerMax and PowerSoil DNA Isolation Kits (MOBIO) replacing the initial
158 bead-beating step by an overnight incubation at 56°C with proteinase K (0.4 mg/ml).
159 Filter samples were extracted with the DNeasy blood & tissue kit (QIAGEN) following the
160 "SX filters without preservation buffer" developed by Spens et al. (2017). Negative
161 controls were included in each batch of DNA extraction and followed the whole process.
162 DNA concentration was measured with the Quant-iT dsDNA HS assay kit (Thermo
163 Scientific) using a Qubit 2.0 Fluorometer (Life Technologies), purity was inferred from
164 260/280 and 260/230 absorbance ratios with the ND-1000 Nanodrop (Thermo
165 Scientific), and integrity was assessed by electrophoresis in 0.7% agarose. Two primer
166 pairs widely used to assess eukaryotic biodiversity (Hebert, Ratnasingham, & de Waard,

167 2003; Amaral-Zettler, McCliment, Ducklow, & Huse, 2009) were used:
168 mlCOIintF/dgHCO2198 ("COI primers"), targeting a 313 bp fragment of the cytochrome
169 oxidase I (COI) gene (Leray et al., 2013) and 1389F/1510R ("18S primers"), targeting a
170 variable length fragment (87 to 186 bp) of the V9 hypervariable region of the 18S rRNA
171 gene (Amaral-Zettler et al., 2009). DNA extracted from zooplankton, fouling organism
172 and filter samples were amplified with both primer pairs, whereas DNA from sediment
173 was only amplified with the "COI primers". PCR amplifications were performed in two
174 rounds. For the first PCR, 1 µl of genomic DNA (5 ng/µl) was added to a mix consisting in
175 5 µl of 1X Phusion Master Mix (Thermo Scientific), 0.2 µl of each primer (0.2 µM) and
176 3.6 µl of MilliQ water. For the "18S primers", PCR conditions consisted on an initial 3 min
177 denaturation step at 98 °C; followed by 25 cycles of 10 s at 98 °C, 30 s at 57 °C and 30 s
178 at 72 °C and finally 10 min at 72 °C. For the "COI primers", PCR conditions consisted on
179 an initial 3 min denaturation step at 98 °C; followed by 35 cycles of 10 s at 98 °C, 30 s at
180 46 °C and 45 s at 72 °C and finally 5 min at 72 °C. Negative controls were included within
181 each set of PCRs. For each DNA extract, three PCR amplifications were performed and
182 pooled. Once purified using AMPure XP beads (Beckman Coulter), the mixed PCR
183 products were used as template for the generation of dual-indexed amplicons in a
184 second PCR round following the "16S Metagenomic Sequence Library Preparation"
185 protocol (Illumina) using the Nextera XT Index Kit (Illumina). Multiplexed PCR products
186 were purified again using the AMPure XP beads, quantified using Quant-iT dsDNA HS
187 assay kit and a Qubit 2.0 Fluorometer (Life Technologies), normalized to equal
188 concentration and sequenced using the 2 x 300 paired-end MiSeq (Illumina). Reads were
189 demultiplexed based on their barcode sequences.
190

191 *Raw read preprocessing, clustering and taxonomic assignment*

192 After quality checking of demultiplexed paired-end reads with FASTQC (Andrews, 2010),
193 forward and reverse primers were removed using Cutadapt (Martin, 2011) with the
194 "anchored 5' adapter" and for "paired end reads" options and with the "linked adapter"
195 option for COI and 18S respectively. Forward and reverse reads were merged using PEAR
196 (Zhang, Kobert, Flouri, & Stamatakis, 2014) with a minimum sequence overlap of 217 bp
197 and maximum amplicon length of 313 bp and of 80 and 190 bp for COI and 18S barcodes
198 respectively. Merged reads with average Phred quality score lower than 20 were
199 removed with Trimmomatic (Bolger, Lohse, & Usadel, 2014). Using Mothur (Schloss et
200 al., 2009), sequences without ambiguous bases were aligned to BOLD
201 (https://www.boldsystems.org) or SILVA (https://www.arb-
202 silva.de/documentation/release-132/) for COI and 18S respectively, and only those
203 covering the barcode region were kept. Chimeras, detected using *de novo* mode of
204 UCHIME (Edgar, Haas, Clemente, Quince, & Knight, 2011), were removed, and
205 remaining reads were clustered into OTUs using Swarm 2.2.1 (Mahé, Rognes, Quince,
206 de Vargas, & Dunthorn, 2014). "Singleton" OTUs, composed by a single read, were
207 removed, and the remaining OTUs were taxonomically assigned using the Naïve Bayes
208 Classifier (Wang, Garrity, Tiedje, & Cole, 2007) using BOLD (accessed in May 2018) or
209 PR2 (release 4.10.0) databases for COI and 18S respectively.

210

211 *Community analyses*

212 Apart from the complete dataset, we created two subsets: a taxa targeted through PBBS
213 dataset, including only reads classified to the class level and excluding those matching
214 to non-targeted groups for PBBS such as Mammalia, Aves, Insecta, Collembola,

215 Arachnida and all classes of Fungi; and a NICS dataset, including only reads matching
216 either the Non-Indigenous and Cryptogenic Species (NICS) previously detected in the
217 port of Bilbao (Martínez & Adarraga, 2006; Adarraga & Martínez, 2011; Butrón, Orive, &
218 Madariaga, 2011; Adarraga & Martínez, 2012; Zorita et al., 2013; Tajadura, Bustamante,
219 & Salinas, 2016) or the AquaNIS database (AquaNIS. Editorial board., 2015). Most
220 analyses were conducted using RStudio (Team, 2015) with *vegan* (Oksanen et al., 2013),
221 *adespatial* (Dray et al., 2018) and *indicspecies* (De Cáceres & Legendre, 2009) libraries.
222 Indicator species analyses (Dufrene & Legendre, 1997) were performed on 1) the taxa
223 targeted through PBBS dataset to identify indicator taxa of each sampling method and
224 2) the NICS dataset to identify non-indigenous indicator taxa of each sampling site.
225 These analyses were based on the IndVal index calculated as the product of the degree
226 of specificity (measuring the uniqueness to a sampling method or site) and the degree
227 of fidelity (measuring the frequency of occurrence within a sampling method or site) of
228 an OTU to a given sampling condition. Statistical significance of associations was
229 assessed by performing 10,000 permutations. The effects of season and locality on taxa
230 targeted through PBBS communities were tested for significance using a permutational
231 multivariate analysis of variance (PERMANOVA) after checking for multivariate
232 homogeneity of group dispersions (betadisper). PERMANOVA and betadisper were
233 performed on OTU abundance (Hellinger distance) and presence/absence (Jaccard
234 dissimilarities). The contribution of replacement (changes in OTU identity) and
235 nestedness (richness differences where one sample is a subset of a richer sample) to
236 beta diversity of taxa targeted through PBBS between seasons and between sites was
237 computed using the relativized nestedness index of Podani and Schmera (2011) based
238 on Jaccard dissimilarities matrix. For each sampling method, replacement and

239 nestedness were calculated between pairwise comparisons of 1) samples belonging to
240 the same site but sampled at different seasons (season variation of beta diversity), and
241 2) samples belonging to the same season but sampled at different sites (spatial variation
242 of beta diversity). The mean proportion of nestedness and replacement contribution to
243 beta-diversity between sites and between seasons was then calculated.
244
245 **Results**
246 *Communities retrieved by each sampling method and genetic marker*
247 A total of 5,718,639 and 7,055,675 COI and 18S barcode reads were kept for
248 analysis (Table S1). Negative controls showed little contamination (with only 19, 36 and
249 55 of COI and 31 and 741 18S reads belonging to taxa targeted through PBBS). The reads
250 corresponding to the 192 samples collected at the port of Bilbao resulted in 40,318 and
251 20,473 OTUs for COI and 18S respectively. For all sampling methods, rarefaction curves
252 of OTUs against reads approached saturation, indicating enough sequencing effort (Figs
253 2A and S2). The majority of the OTUs, 89% for COI and 73% for 18S, were unique to one
254 sampling method (Fig. 2B), and, for both barcodes, but particularly for 18S, filtered
255 water (aimed at retrieving eDNA from macroorganisms and microbial eukaryotes)
256 resulted in higher OTU richness and unique OTUs than other sampling methods (Figs 2A,
257 B and S2). Different sampling methods retrieved distinct biological communities as
258 observed with the PCA (Fig. 2C) and confirmed with PERMANOVA analyses (COI:
259 R^2=0.28, p-value= 9.99e-05; 18S: R^2=0.30, p-value=9.99e-05).
260 The percentage of reads assigned to species or higher taxonomic levels differed
261 between sampling methods and barcodes. The majority of the reads were assigned to
262 species level for zooplankton nets (COI: 84% and 18S: 75%), settlement plates (COI: 69%

263 and 18S: 69%) and sediment grabs (COI: 64%); for filtered water, although 51% of reads
264 were assigned to species level with 18S, only 8% could be assigned to a species with COI
265 (Fig. 3A). Filtered water analyzed with COI had also the largest proportion of reads that
266 could not be assigned to phylum (87%). This is possibly due to the amplification of non-
267 metazoan taxa, which are underrepresented in the BOLD database (Fig. S3), which was
268 confirmed with BLAST searches against GenBank (data not shown).

269 Both barcodes detected a wide range of eukaryotic groups (Fig. 3B). With COI,
270 the greatest majority of reads belonged to metazoans, with Arthropoda dominating all
271 sampling methods except sediment grabs, dominated by Annelida. With 18S, a more
272 diverse spectrum of taxa was retrieved, including phytoplankton and macroalgae; as
273 expected, the number of reads assigned to these phyla in zooplankton nets and
274 settlement plates was very low in comparison to filtered water.

275

276 *Distribution of taxa targeted through port biological baseline survey*

277 A total of 16,828 and 9,091 OTUs were kept as taxa targeted through PBBS for
278 COI and 18S respectively. From those, indicator analysis identified respectively 2,600
279 (15%) and 1,700 (19%) OTUs significantly associated to one of the sampling methods.
280 Settlement plates were associated with 1,268 (COI) and 808 (18S) OTUs, zooplankton
281 nets with 1,001 (COI) and 315 (18S) OTUs, and sediment grabs with 238 (COI) OTUs. Only
282 12 OTUs (9 of which were metazoans) were indicators of filtered water with COI while
283 580 (3 of which were metazoans), were associated to this sampling method with 18S.
284 We observed expected strong associations of some taxa to one particular sampling
285 method: fouling bivalves (Ostreoida, Mytilida) with settlement plates, copepods
286 (Calanoida, Cyclopoida) with zooplankton nets and sea urchins (Spatangoida) with

287 sediment grabs (Fig. 4). Yet, less obvious associations were also observed such as
288 barnacles (Sessilia) and two polychaeta orders (Spionida, Phyllodocida) with
289 zooplankton nets or dinoflagellates (Peridiniales and Gymnodiniales) with settlement
290 plates, illustrating the advantages of a complementary sampling approach to recover
291 the diversity of these taxonomic groups.

292 The need for complementarity of sampling methods was further confirmed by
293 PCA on PBBS target taxa (Fig. 5). For both barcodes, expected taxonomic differences
294 were found between the sampling methods: sediment grabs were characterized by the
295 polychaeta *Maldane glebifex* (Fig. 5B), zooplankton nets were distinguished by several
296 copepods (Fig. 5B, D) and settlement plates, by the encrusting *Semibalanus balanoides*
297 (Fig. 5D). Filtered water communities were characterized differently according to each
298 barcode. For COI, filtered water samples were not differentiated as a distinct group and
299 were in general close to those retrieved with zooplankton samples collected at the same
300 season (Fig. 5A). In contrast, for 18S, filtered water was different due to the presence of
301 the phytoplankton species *Phaeocystis globosa* (Fig. 5D); yet, when targeting only
302 metazoan taxa, the patterns observed with COI were similarly to the ones retrieved with
303 18S (Fig. S4A). Concerning filtered samples, when looking only at metazoan taxa, the
304 proportion of OTUs detected was largely different compared to the other methods (Fig.
305 S5). For instance, Decapoda orders had a low OTU diversity with filtered water (COI:
306 0,7%; 18S: 7%), whereas more OTU diversity could be detected for Calanoida (COI: 8%;
307 18S: 35%), Sabellida (COI: 15%; 18S: 28%) and Leptothecata (COI: 13%; 18S: 58%). In
308 general, filtered water recovered the smallest metazoan OTU diversity.

309

310 *Influence of sampling seasonality and locality on detected biodiversity*

311 Seasonal variations contributed significantly to differences in community
312 composition of zooplankton nets and filtered water but not of sediment grabs (Table
313 S2). Total OTU richness and unique OTU richness varied between seasons for all
314 sampling methods (Fig. 6A, B). Spring and late summer were generally richer than
315 autumn and winter, excepted for filtered water communities depicted with 18S. For
316 sediment grabs and zooplankton nets with 18S, a strong proportion of OTUs found in
317 autumn and winter were a subset of OTUs retrieved in late summer and spring. Indeed,
318 the nestedness component of sediment grab assemblages represented between 60 to
319 70% of total compositional variations among these pairs of seasons (Fig. 6D). For
320 zooplankton net assemblages, 63 and 72% of nestedness was observed between late
321 summer and spring and between late summer and autumn respectively, suggesting that
322 late summer recovered a majority of spring and autumn diversity (Fig. 6D). Thus, the
323 combination of spring and late summer for sediment grabs retrieved 85% of the total
324 OTU richness over the four seasons, while for zooplankton nets, in late summer alone
325 71% was retrieved (Fig. 6C, E). For filtered water and settlement plates, seasonal
326 community changes were driven by both OTU replacement and nestedness, with similar
327 relative contributions (Fig. 6D). The seasonal influence observed in intra-port samples
328 was also observed between ports (Fig. S6). The differences in communities between
329 Bilbao, A Coruña and Vigo (ports belonging to the same ecoregion but separated by over
330 500 km) during the same season was generally smaller than that between seasons in the
331 same port, indicating that communities were driven by seasonality rather than location
332 (Fig. S6 and Table S3). This pattern was more pronounced with 18S than with COI.

333 Locality within the Bilbao port appeared to impact on benthic assemblages (Table
334 S2), since sites outside the estuary (sites 1 to 3) were different from site 4 inside the
335 estuary; mainly, the polychaete *Maldane glebifex* was less abundant in site 4 (Fig. 5B).
336 Regarding zooplankton nets, site 4 was the main driver of difference in communities as,
337 when not considered, no significant differences between sites were observed (Table S2).
338 Each site harbored a similar proportion of the total OTU richness found by each sampling
339 method (Fig. 7C). This proportion did not exceed 60%. Indeed, OTU replacement
340 contributed more to community variation among sites than nestedness, especially with
341 sediment grab and settlement plates (Fig. 7D). OTU replacement was more important
342 when comparing sites from outside the estuary (sites 1, 2 and S) with site 4 inside the
343 estuary, while it contributed less to community variation among sites 1,2 and 3. This is
344 congruent with site 4 having generally more unique OTUs than the other sites (Fig. 7B).
345 An exception was observed for filtered water with COI, where site 4 had the lowest OTU
346 richness and unique OTU richness, and where nestedness contributed more to the
347 variation in community composition in comparison to sites 1 and 2 (Fig. 7A, B, D).
348
349 *Detection of Non-Indigenous and Cryptogenic Species (NICS)*
350 Our port baseline biodiversity survey detected 79 NICS, among which 29 were
351 previously recorded in the port Bilbao (Table S4). Most of the other species (43) were
352 previously detected NICS inside the port's Large Marine Ecoregion (LME) but, for 7 of
353 them (*Ammothea hilgendorfi, Bugulina fulva, Grandidierella japonica, Melita nitida,*
354 *Neodexiospira brasiliensis, Pseudochattonella verruculosa, Tubificoides pseudogaster*),
355 it was, to our knowledge, the first report in the LME. The indicator species analysis
356 performed on non-indigenous OTU revealed that site 4 was associated to the highest

357 non-indigenous and cryptogenic diversity with, for COI, 603 OTU indicators, and for 18S,
358 24 OTU indicators. These OTUs belonged to 10 NICS with *Oithona davisae*, *Acartia tonsa*
359 and *Allita succinea*, found by both barcodes, *Ficopomatus enigmaticus,* with 18S and
360 *Amphibalanus eburneus*, *Grandidierella japonica*, *Polydora cornuta*, *Austrominius*
361 *modestus*, *Monocorophium acherusicum*, *Xenostrobus securis,* with COI. Regarding the
362 three other sites, for COI, site 1 was associated to one *Clytia hemisphaerica* OTU and to
363 one *Balanus trigonus* OTU, site 2, to three *Clytia hemisphaerica* and four *Mytilus edulis*
364 OTUs and site 3, to one *Balanus trigonus* OTU. For 18S, only site 1 was associated to one
365 OTU, assigned to *Gymnodinium aureolum.*
366
367 **Discussion**
368 In order to be comprehensive, port baseline surveys need to have an extensive spatial
369 and temporal coverage so that most biodiversity is captured by the collected samples.
370 Unlike morphological taxonomic identification, metabarcoding allows the analysis of
371 hundreds of samples is a cost-effective manner (Aylagas, Borja, Muxika, & Rodríguez-
372 Ezpeleta, 2018), enabling for comprehensive PBBS. Metabarcoding-based PBBS can
373 follow a taxonomy dependent or independent approach. The first approach requires
374 comparing the sequences against reference databases for taxonomic assignment,
375 whose accuracy strongly depends on the completeness of the reference database; this
376 reference database is particularly relevant for detection of NICS. The second approach
377 consists on inferring diversity metrics from sequences to compare across samples for
378 which the development of standardized procedures is mandatory so that comparisons
379 are possible. Here we have focused on developing such procedures by evaluating
380 alternative strategies.

381

382 *A comprehensive metabarcoding-based port baseline survey should rely on a*
383 *combination of sampling methods*

384 Most of the past attempts on using metabarcoding for port monitoring have
385 relied on a unique sampling method (Brown et al., 2016; Zaiko et al., 2016; Borrell et al.,
386 2017; Grey et al., 2018; Lacoursière-Roussel et al., 2018) and only one evaluated the
387 importance of sampling methods (Koziol et al., 2019). Here, in order to define
388 comprehensive guidelines for developing standardizable port sampling protocols, we
389 have adapted the HELCOM/OSPAR guidelines to collect zooplankton, fouling, and
390 sediment samples for genetic analyses. In agreement with previous work (Koziol et al.,
391 2019), our metabarcoding analysis of the collected samples has shown that each
392 sampling method recovered a distinct subset of the port community, and that, despite
393 some taxa being expectedly associated to a given sampling method, the total diversity,
394 even within specific taxonomic groups, was only recovered by combining different
395 methods. Thus, combining sampling methods is needed to maximize biodiversity
396 detection.

397

398 *Metabarcoding of eDNA from water samples using universal primers does not provide a*
399 *comprehensive view of the overall macroorganismal diversity*

400 As an attempt to find a sampling method that would capture a large fraction of
401 the diversity with limited effort, we also explored metabarcoding on eDNA extracted
402 from filtered water samples. Recent applications of this technique have shown its
403 potential to recover spatial and temporal variation of macroorganismal communities
404 (Bista et al., 2017; O'Donnell et al., 2017), including detection of non-indigenous taxa

405 (Klymus, Marshall, & Stepien, 2017). Consequently, recent studies have considered
406 eDNA for port surveys (Borrell et al., 2017; Grey et al., 2018; Lacoursière-Roussel et al.,
407 2018). Yet, as suggested by other studies in coral reef sites (Djurhuus et al., 2018) and
408 freshwater streams (Macher et al., 2018) and recently in ports (Koziol et al., 2019), our
409 analyses show that, compared to bulk sample metabarcoding, eDNA metabarcoding
410 recovered only a subset of the metazoan diversity and did not provide additional
411 information on targeted groups. This suggests that, despite requiring less sampling
412 effort, COI and 18S based metabarcoding of eDNA obtained from filtered water cannot
413 be used to replace the need of obtaining bulk samples (using zooplankton nets, grabs
414 and settlement plates) for PBBS taxa detection. In this context, at least two alternatives
415 are possible to improve eDNA-based biodiversity assessments: i) increasing sequencing
416 depth (Grey et al., 2018) or ii) using group specific markers (Koziol et al., 2019). Yet, none
417 of them ensures full biodiversity recovery and both significantly increase costs. Thus, the
418 decision between increasing sampling depth and/or using multiple group specific
419 primers for eDNA and adding multiple substrate sampling needs to be carefully
420 considered and will be an important area of future research to optimize metabarcoding-
421 based port monitoring.
422
423 *A comprehensive metabarcoding-based port baseline survey should include*
424 *spatiotemporal sampling*
425 While it is expected that increased temporal coverage will retrieve more taxa,
426 the costs associated to extensive sampling demands an evaluation of the benefits gained
427 by sampling at each season. Indeed, the HELCOM/OSPAR protocol limits sampling to late
428 summer for sediment and fouling, and to spring and late summer for plankton. Here we

429 show that for sediment and fouling, spring sampling provides higher diversity than late
430 summer. Although sampling in late summer could be appropriate for morphological
431 taxonomy because of the more abundance of adult individuals, for metabarcoding,
432 sampling in spring is preferable because during this season i) sizes of organisms are less
433 variable and thus metabarcoding is less likely to under detect small organisms (Elbrecht,
434 Peinert, & Leese, 2017) and ii) species diversity is at its maxima due to being a high
435 recruitment period with abundant organisms at early life stages (Bijleveld et al., 2018).
436 For zooplankton diversity, our results show that the HELCOM/OSPAR approach of
437 sampling in spring and late summer produces the highest diversity. However, the data
438 obtained from filtered water were inconclusive. Although spring was the most diverse
439 season for COI, this was not the case for 18S, and, overall, other combinations of seasons
440 appeared to perform better. Previous studies have already shown that seasonal
441 variations are important considerations for eDNA studies (Lacoursière-Roussel et al.,
442 2018), but further characterizations are needed over multiple years to design an
443 adequate protocol for maximizing biodiversity recovery with eDNA. Our study also
444 highlights the influence of sampling site in recovered biodiversity, although only one of
445 the four sites sampled, the less saline one located at the inner part of the estuary,
446 showed a significantly different community composition.

447 Concerning spatial sampling, we observed no significant differences between
448 sites when using eDNA, except between the only site located in the inner estuary, and
449 the others and only when using zooplankton, and significant differences overall when
450 using sediment grabs. The proportion of the total OTUs found in a single site was low,
451 suggesting that spatially comprehensive sampling is crucial to recover the port´s
452 biodiversity. Interestingly, site 4 was not only the most different from all four sites but

453 was also the one that recorded the largest number of NIS. In coherence with our
454 findings, it has been observed that brackish environments favors NIS settlement (Zorita
455 et al., 2013) because these species usually support wider range of salinity (Cardeccia et
456 al., 2018). Thus, samples for port monitoring should include those with a wide range of
457 abiotic conditions and covering priority sampling, such as highly active ship berths,
458 potential reservoirs of newly arrived NIS (Hewitt & Martin, 2001).
459
460 *Metabarcoding provides valuable information on Non-Indigenous Species*
461 Our metabarcoding port biodiversity baseline survey detects species previously
462 recorded in the port of Bilbao and species known to be present in the port´s large marine
463 ecoregion. Importantly, it also detects presence of seven species for which no records
464 exist so far in this ecoregion, highlighting the potential of metabarcoding for early NICS
465 detection. Nonetheless, not all NICS species previously recorded in Bilbao were found
466 by our analyses. This might be due to these species not being present in the port at the
467 time of our survey, to biases of the metabarcoding process such as differential DNA
468 extraction, primer non-specificity, etc (Xiong, Li, & Zhan, 2016), or to database
469 incompleteness. Primer bias and/or database incompleteness could also explain why 35
470 and 42% of NICS were uniquely detected with COI and 18S respectively. Yet, from the
471 total species included in the AquaNIS database (n=1083), only 460 and 369 are included
472 in the BOLD and PR2 databases respectively. This stresses the need of increasing
473 reference databases and of using multiple universal primers for species detection (Grey
474 et al., 2018). Importantly, our study confirmed that metabarcoding can detect species
475 occurring at low abundance (Pochon, Bott, Smith, & Wood, 2013) as we found the non-
476 indigenous amphipod *Melita nitida* with only 18 reads while it took intensive surveys in

477 2013, 2014 and 2016 in 3 distinct sampling regions of the Bay of Biscay to record 76
478 individuals (Gouillieux, Lavesque, Blanchet, & Bachelet, 2016).
479
480 *Guidelines and recommendations for standardized and comparable metabarcoding-*
481 *based port baseline surveys*
482 Based on the comparative analysis of samples collected using alternative sampling
483 methods, at different seasons of the year and from different port locations, we have
484 determined the effect of each factor on derived biodiversity estimates. This has allowed
485 to define a series of standard guidelines that maximize recovered biodiversity while
486 minimizing sampling effort and which can be applied to ensure comparability among
487 ports:
488 1. Sampling site selection: Obtain information about port traffic and potentially
489 relevant abiotic conditions (*i.e,* salinity gradient) and select at least three sites,
490 including active and inactive berths and covering the salinity range (*i.e.* marine
491 to brackish).
492 2. Sampling method selection: Collect samples from, at least, zooplankton nets,
493 sediment grabs, filtered water and settlement plates (see Material and Methods
494 for detailed sampling techniques).
495 3. Sampling season selection: In temperate latitudes, collect benthic (sediment
496 grabs) and fouling (settlement plates) macrofauna at least in spring and late
497 summer, zooplankton at least in late summer and eDNA (from filtered water) at
498 least in spring.
499 4. Metabarcoding data generation: Follow standardized protocols for sample pre-
500 processing, DNA extraction, barcode amplification, using at least two universal

primers, and sequencing (see Material and Methods) and use established procedures for read preprocessing, OTU clustering and taxonomic assignment (see provided analysis code).

5. Non-Indigenous Species detection: Compare the list of species found in the port with comprehensive and updated databases of Non-Indigenous Species (*i.e.* AquaNIS) for a first indication of potential invaders presence.

These guidelines and considerations are particularly relevant due to the recent entry into force of the International Convention for the Control and Management of Ships' Ballast Water and Sediments as they will allow performing standardized risk assessments to evaluate the possibility of granting an exemption for ships travelling between specific ports.

Acknowledgments

This manuscript is a result of the Aquainvad-ED (AQUAtic INVADers: Early Detection, Control and Management - www.aquainvad-ed.com) project funded by the European Union (Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement no 642197). We thank the Port Authority of Bilbao for authorizing sampling in the port and Manuel González for making the necessary arrangements for it. We also thank Iñaki Mendibil and Elisabete Bilbao for laboratory work, Ekaitz Erauskin, Iker Urtizberea and Lander Larrañaga Juarez for sampling, Anders Lanzen for providing comments on the manuscript, and Carlota Bermejo, student of the postgraduate course in Scientific Illustration of the UPV/EHU, for producing Figure 1. This manuscript is contribution X [to be completed upon acceptance] from the Marine Research Division of AZTI.

525 **Authors' contributions**

526 AR, OB and NR-E conceived the study and designed methodology; AR and OB collected
527 the samples; AR preformed laboratory analyses; AR and NR-E analyzed the data; AR and
528 NR-E wrote the manuscript. All authors contributed critically to the drafts and gave final
529 approval for publication.

530

531 **Data Accessibility**

532 Raw demultiplexed MiSeq reads are available at NCBI SRA Bioproject PRJNA515494.
533 Scripts used for OTU table generation and taxonomic assignment are available at
534 https://github.com/rodriguez-ezpeleta/metabarcoding-pbbs.

535

536 **References**

537 Adarraga, I., & Martínez, J. (2011). First record of invasive *Theora lubrica* (Mollusca:
538 Bivalvia: Semelidae) in the Atlantic Ocean. *Marine Biodiversity Records*, *4*.
539 doi:10.1017/S1755267211001011

540 Adarraga, I., & Martínez, J. (2012). First record of the invasive brackish water mytilid
541 *Limnoperna securis* (Lamarck, 1819) in the Bay of Biscay. *Aquatic Invasions*, *7*(2).

542 Amaral-Zettler, L. A., McCliment, E. A., Ducklow, H. W., & Huse, S. M. (2009). A Method
543 for Studying Protistan Diversity Using Massively Parallel Sequencing of V9
544 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. *PLOS ONE*, *4*(7),
545 e6372. doi:10.1371/journal.pone.0006372

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

AquaNIS. Editorial board. (2015). Information system on Aquatic Non-Indigenous and Cryptogenic Species. World Wide Web electronic publication. www.corpi.ku.lt/databases/aquanis. Version 2.36+.

Awad, A., Haag, F., Anil, A. C., & Abdulla, A. (2014). *Guidance on Port Biological Baseline Surveys (PBBS)*. GEF-UNDP-IMO GloBallast Partnerships, IOI, NIO, IUCN.

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. doi:10.1016/j.ecolind.2018.07.044

Aylagas, E., Mendibil, I., Borja, Á., & Rodríguez-Ezpeleta, N. (2016). Marine Sediment Sample Pre-processing for Macroinvertebrates Metabarcoding: Mechanical Enrichment and Homogenization. *Frontiers in Marine Science*, *3*. doi:10.3389/fmars.2016.00203

Bax, N., Williamson, A., Aguero, M., Gonzalez, E., & Geeves, W. (2003). Marine invasive alien species: a threat to global biodiversity. *Marine Policy*, *27*(4), 313–323. doi:10.1016/S0308-597X(03)00041-1

Bijleveld, A. I., Compton, T. J., Klunder, L., Holthuijsen, S., Horn, J. ten, Koolhaas, A., … Veer, H. W. van der. (2018). Presence-absence of marine macrozoobenthos does not generally predict abundance and biomass. *Scientific Reports*, *8*(1), 3039. doi:10.1038/s41598-018-21285-1

Bishop, M. J., & Hutchings, P. A. (2011). How useful are port surveys focused on target pest identification for exotic species management? *Marine Pollution Bulletin*, *62*(1), 36–42. doi:10.1016/j.marpolbul.2010.09.014

570 Bista, I., Carvalho, G. R., Walsh, K., Seymour, M., Hajibabaei, M., Lallias, D., … Creer, S.
571 (2017). Annual time-series analysis of aqueous eDNA reveals ecologically
572 relevant dynamics of lake ecosystem biodiversity. *Nature Communications*, *8*,
573 14087. doi:10.1038/ncomms14087
574 Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for
575 Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.
576 doi:10.1093/bioinformatics/btu170
577 Borrell, Y. J., Miralles, L., Huu, H. D., Mohammed-Geba, K., & Garcia-Vazquez, E. (2017).
578 DNA in a bottle—Rapid metabarcoding survey for early alerts of invasive species
579 in ports. *PLOS ONE*, *12*(9), e0183347. doi:10.1371/journal.pone.0183347
580 Bott, N. J. (2015). Detection of Invasive Species. In S.-K. K. Prof (Ed.), *Springer Handbook*
581 *of Marine Biotechnology* (pp. 441–453). Springer Berlin Heidelberg.
582 doi:10.1007/978-3-642-53971-8_16
583 Brown, E. A., Chain, F. J. J., Zhan, A., MacIsaac, H. J., & Cristescu, M. E. (2016). Early
584 detection of aquatic invaders using metabarcoding reveals a high number of non-
585 indigenous species in Canadian ports. *Diversity and Distributions*, *22*(10), 1045–
586 1059. doi:10.1111/ddi.12465
587 Butrón, A., Orive, E., & Madariaga, I. (2011). Potential risk of harmful algae transport by
588 ballast waters: The case of Bilbao Harbour. *Marine Pollution Bulletin*, *62*(4), 747–
589 757.
590 Cardeccia, A., Marchini, A., Occhipinti-Ambrogi, A., Galil, B., Gollasch, S., Minchin, D., …
591 Ojaveer, H. (2018). Assessing biological invasions in European Seas: Biological
592 traits of the most widespread non-indigenous species. *Estuarine, Coastal and*
593 *Shelf Science*, *201*, 17–28. doi:10.1016/j.ecss.2016.02.014

Carlton, J. T. (1999). The scale and ecological consequences of biological invasions in the World's oceans. (pp. 195–212). Presented at the Population and Community Biology Series Volume 24, Kluwer Academic Publishers.

Comtet, T., Sandionigi, A., Viard, F., & Casiraghi, M. (2015). DNA (meta) barcoding of biological invasions: a powerful tool to elucidate invasion processes and help managing aliens. *Biological Invasions*, *17*(3), 905–922.

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

David, M., Gollasch, S., & Pavliha, M. (2013). Global ballast water management and the 'same location' concept: a clear term or a clear issue? *Ecological Applications*, *23*(2), 331–338. doi:10.1890/12-0992.1

De Cáceres, M. D., & Legendre, P. (2009). Associations between species and groups of sites: indices and statistical inference. *Ecology*, *90*(12), 3566–3574.

Deiner, K., Fronhofer, E. A., Mächler, E., Walser, J.-C., & Altermatt, F. (2016). Environmental DNA reveals that rivers are conveyer belts of biodiversity information. *Nature Communications*, *7*, 12544. doi:10.1038/ncomms12544

Djurhuus, A., Pitz, K., Sawaya, N. A., Rojas-Márquez, J., Michaud, B., Montes, E., … Breitbart, M. (2018). Evaluation of marine zooplankton community structure through environmental DNA metabarcoding: Metabarcoding zooplankton from eDNA. *Limnology and Oceanography: Methods*, *16*(4), 209–221. doi:10.1002/lom3.10237

Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., … Wagner, H. H. (2018). *adespatial: Multivariate Multiscale Spatial Analysis*. Retrieved from https://CRAN.R-project.org/package=adespatial

Dufrene, M., & Legendre, P. (1997). Species Assemblages and Indicator Species: The Need for a Flexible Asymmetrical Approach. *Ecological Monographs*, *67*(3), 345. doi:10.2307/2963459

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, *27*(16), 2194–2200. doi:10.1093/bioinformatics/btr381

Elbrecht, V., Peinert, B., & Leese, F. (2017). Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. *Ecology and Evolution*, *7*(17), 6918–6926. doi:10.1002/ece3.3192

Elbrecht, V., Vamos, E. E., Meissner, K., Aroviita, J., & Leese, F. (2017). Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. *Methods in Ecology and Evolution*, *8*(10), 1265–1275. doi:10.1111/2041-210X.12789

Gittenberger, A., Rensing, M., Niemantsverdriet, P., Schrieken, N., & Stegenga, H. (2014, November). Port of Rotterdam survey and monitoring non-native species conform HELCOM/OSPAR protocol.

Gouillieux, B., Lavesque, N., Blanchet, H., & Bachelet, G. (2016). First record of the non-indigenous *Melita nitida Smith*, 1873 (Crustacea: Amphipoda: Melitidae) in the Bay of Biscay (NE Atlantic). *BioInvasions Records*, *5*(2), 85–92.

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

641 environmental DNA metabarcoding surveys. *Scientific Reports*, *8*(1).
642 doi:10.1038/s41598-018-27048-2
643 Hebert, P. D., Ratnasingham, S., & de Waard, J. R. (2003). Barcoding animal life:
644 cytochrome c oxidase subunit 1 divergences among closely related species.
645 *Proceedings of the Royal Society of London B: Biological Sciences*, *270*(Suppl 1),
646 S96–S99.
647 HELCOM/OSPAR. (2013). Joint HELCOM/OSPAR Guidelines for the Contracting Parties of
648 OSPAR and HELCOM on the granting of exemptions under International
649 Convention for the Control and Management of Ships' Ballast Water and
650 Sediments, Regulation A-4 (OSPAR Agreement 2013-09), 45 pp.
651 Hewitt, C. L., & Martin, R. B. (2001). Revised protocols for baseline port surveys for
652 introduced marine species–design considerations, sampling protocols and
653 taxonomic sufficiency. *Centre for Research on Introduced Marine Pests Technical*
654 *Report*, (22).
655 IMO. (2004). International Convention for the Control and Management of Ships' Ballast
656 Water and Sediments, 2004. International Maritime Organization, 13 Feb 2004,
657 London.
658 Jeunen, G.-J., Knapp, M., Spencer, H. G., Lamare, M. D., Taylor, H. R., Stat, M., …
659 Gemmell, N. J. (2019). Environmental DNA (eDNA) metabarcoding reveals strong
660 discrimination among diverse marine habitats connected by water movement.
661 *Molecular Ecology Resources*, *19*(2), 426–438. doi:10.1111/1755-0998.12982
662 Klymus, K. E., Marshall, N. T., & Stepien, C. A. (2017). Environmental DNA (eDNA)
663 metabarcoding assays to detect invasive invertebrate species in the Great Lakes.
664 *PLOS ONE*, *12*(5), e0177643. doi:10.1371/journal.pone.0177643

Koziol, A., Stat, M., Simpson, T., Jarman, S., DiBattista, J. D., Harvey, E. S., … Bunce, M. (2019). Environmental DNA metabarcoding studies are critically affected by substrate selection. *Molecular Ecology Resources*, *19*(2), 366–376. doi:10.1111/1755-0998.12971

Kraus, R., Ninčević-Gladan, Ž., Auriemma, R., Bastianini, M., Bolognini, L., Cabrini, M., … Žuljević, A. (2018). Strategy of port baseline surveys (PBS) in the Adriatic Sea. *Marine Pollution Bulletin*. doi:10.1016/j.marpolbul.2018.08.067

Lacoursière-Roussel, A., Howland, K., Normandeau, E., Grey, E. K., Archambault, P., Deiner, K., … Bernatchez, L. (2018). eDNA metabarcoding as a new surveillance approach for coastal Arctic biodiversity. *Ecology and Evolution*. doi:10.1002/ece3.4213

Lehtiniemi, M., Ojaveer, H., David, M., Galil, B., Gollasch, S., McKenzie, C., … Pederson, J. (2015). Dose of truth—Monitoring marine non-indigenous species to serve legislative requirements. *Marine Policy*, *54*, 26–35. doi:10.1016/j.marpol.2014.12.015

Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., … 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. doi:10.1186/1742-9994-10-34

Macher, J.-N., Vivancos, A., Piggott, J. J., Centeno, F. C., Matthaei, C. D., & Leese, F. (2018). Comparison of environmental DNA and bulk-sample metabarcoding using highly degenerate cytochrome c oxidase I primers. *Molecular Ecology Resources*. doi:10.1111/1755-0998.12940

Mahé, F., Rognes, T., Quince, C., de Vargas, C., & Dunthorn, M. (2014). Swarm: robust and fast clustering method for amplicon-based studies. *PeerJ*, *2*, e593. doi:10.7717/peerj.593

Mandelik, Y., Roll, U., & Fleischer, A. (2010). Cost-efficiency of biodiversity indicators for Mediterranean ecosystems and the effects of socio-economic factors. *Journal of Applied Ecology*, *47*(6), 1179–1188. doi:10.1111/j.1365-2664.2010.01864.x

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*(1), 10–12. doi:10.14806/ej.17.1.200

Martínez, J., & Adarraga, I. (2006). Programa de vigilancia y control de la introducción de especies invasoras en los ecosistemas litorales de la costa vasca. 1. *Costa de Gipuzkoa*, *2*.

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. doi:10.7717/peerj.3044

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R. B., … Wagner, H. (2013). Package 'vegan'. *Community Ecology Package, Version*, *2*(9).

Pochon, X., Bott, N. J., Smith, K. F., & Wood, S. A. (2013). Evaluating detection limits of next-generation sequencing for the surveillance and monitoring of international marine pests. *PLoS One*, *8*(9), e73935.

Podani, J., & Schmera, D. (2011). A new conceptual and methodological framework for exploring and explaining pattern in presence - absence data. *Oikos*, *120*(11), 1625–1638. doi:10.1111/j.1600-0706.2011.19451.x

Pyšek, P., Hulme, P. E., Meyerson, L. A., Smith, G. F., Boatwright, J. S., Crouch, N. R., … Wilson, J. R. U. (2013). Hitting the right target: taxonomic challenges for, and of, plant invasions. *AoB PLANTS*, *5*. doi:10.1093/aobpla/plt042

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., … Weber, C. F. (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology*, *75*(23), 7537–7541. doi:10.1128/AEM.01541-09

Seebens, H., Gastner, M. T., & Blasius, B. (2013). The risk of marine bioinvasion caused by global shipping. *Ecology Letters*, *16*(6), 782–790. doi:10.1111/ele.12111

Shaw, J. L. A., Weyrich, L., & Cooper, A. (2016). Using environmental (e)DNA sequencing for aquatic biodiversity surveys: a beginner's guide. *Marine and Freshwater Research*. doi:10.1071/MF15361

Spens, J., Evans, A. R., Halfmaerten, D., Knudsen, S. W., Sengupta, M. E., Mak, S. S. T., … Hellström, M. (2017). 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*(5), 635–645. doi:10.1111/2041-210X.12683

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*(8), 2045–2050.

Tajadura, J., Bustamante, M., & Salinas, J. I. S. (2016). Temporal and spatial variability in the distribution of non-indigenous zoobenthic species in the intertidal hard substrates in the 'Abra of Bilbao' Poster ISOBAY conference 2016

https://www.ehu.eus/documents/5019180/0/Book-of-abstract-ISOBAY15.pdf. *ISOBAY Conference*.

Team, Rs. (2015). RStudio: integrated development for R. *RStudio, Inc., Boston, MA URL Http://Www. Rstudio. Com*, *42*.

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, *73*(16), 5261–5267. doi:10.1128/AEM.00062-07

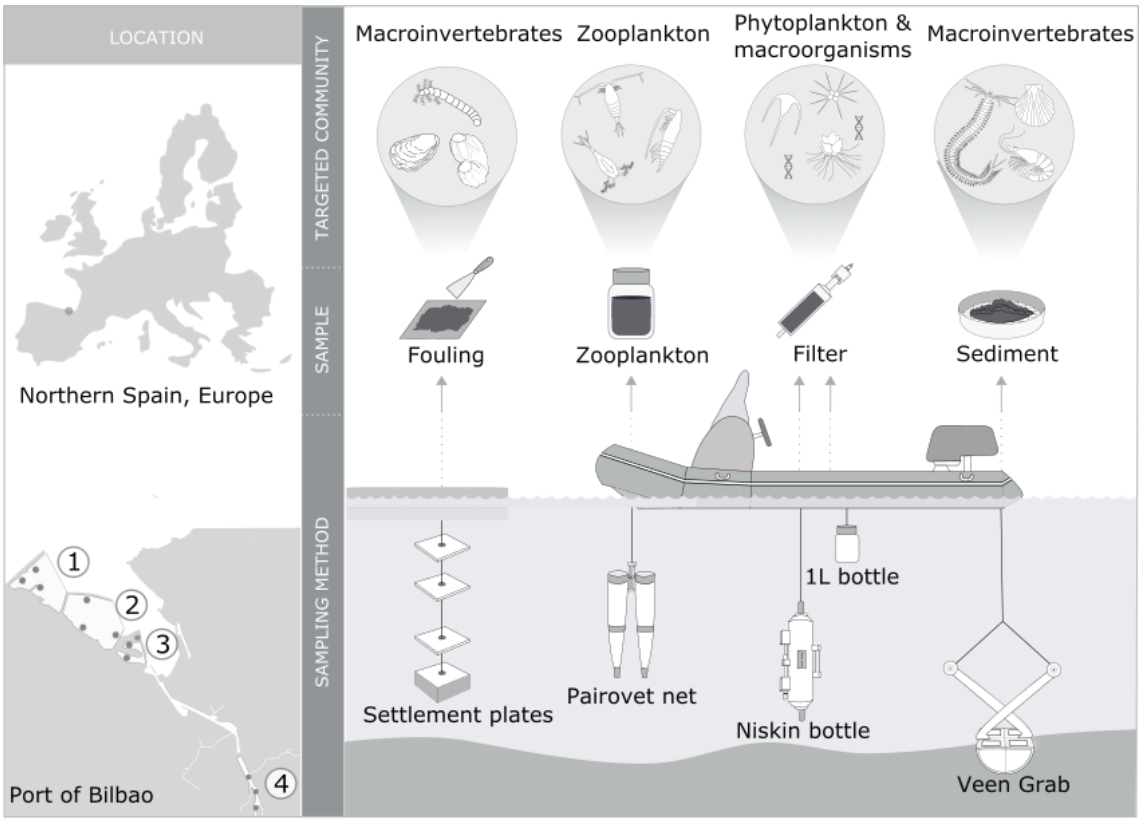
Xiong, W., Li, H., & Zhan, A. (2016). Early detection of invasive species in marine ecosystems using high-throughput sequencing: technical challenges and possible solutions. *Marine Biology*, *163*(6), 1–12.

Zaiko, A., Schimanski, K., Pochon, X., Hopkins, G. A., Goldstien, S., Floerl, O., & Wood, S. A. (2016). Metabarcoding improves detection of eukaryotes from early biofouling communities: implications for pest monitoring and pathway management. *Biofouling*, *32*(6), 671–684. doi:10.1080/08927014.2016.1186165

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, *30*(5), 614–620. doi:10.1093/bioinformatics/btt593

Zorita, I., Solaun, O., Borja, A., Franco, J., Muxika, I., & Pascual, M. (2013). Spatial distribution and temporal trends of soft-bottom marine benthic alien species collected during the period 1989–2008 in the Nervión estuary (southeastern Bay of Biscay). *Journal of Sea Research*, *83*, 104–110.

758 **Figures**



759

760 Figure 1. Sampling protocol depicting sampling location (Port of Bilbao), sites (1, 2, 3 and

761 4) and points per site (dots); and illustrating the sampling methods used and the

762 targeted biological communities.

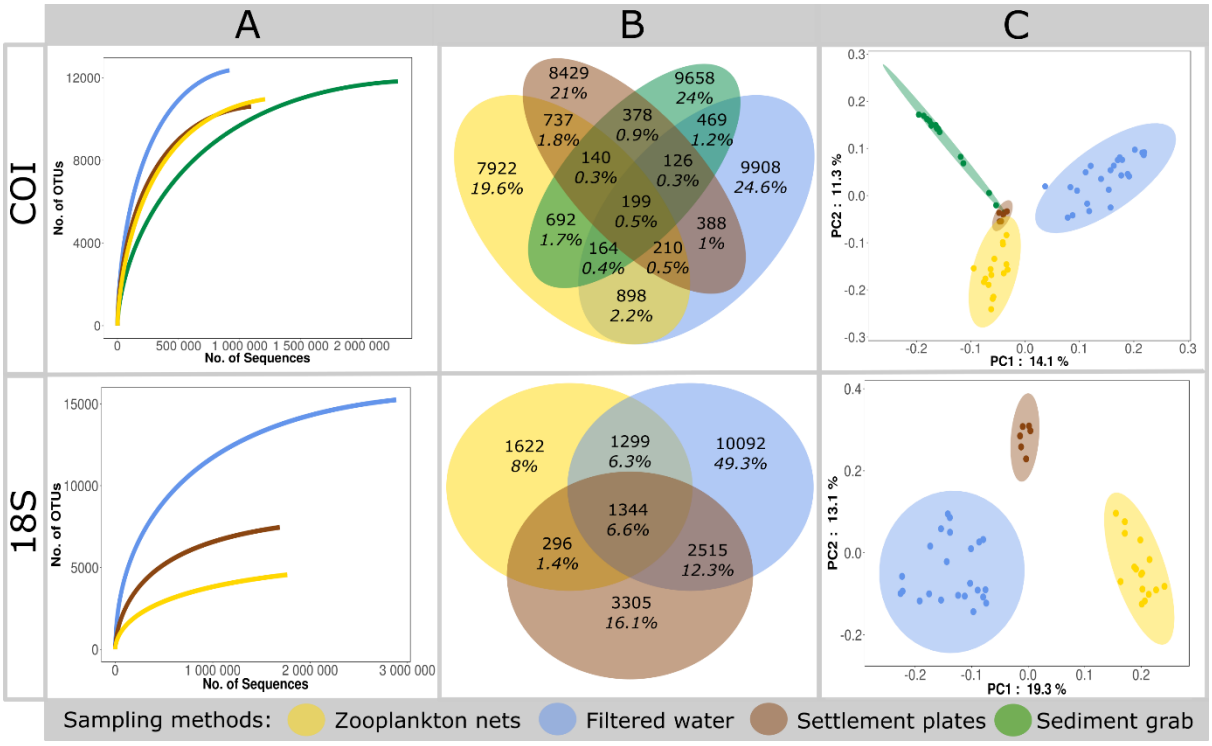


Figure 2. Overall description of community detected per barcode. A: OTU accumulation curves per sampling method. B: Venn diagrams of the number and percentage of OTUs shared between sampling methods. C: Principal Component Analyses of the Hellinger-transformed abundances of OTUs. Sample scores are displayed in scaling 1 with ellipses representing the 95% confidence dispersion of each sampling method.

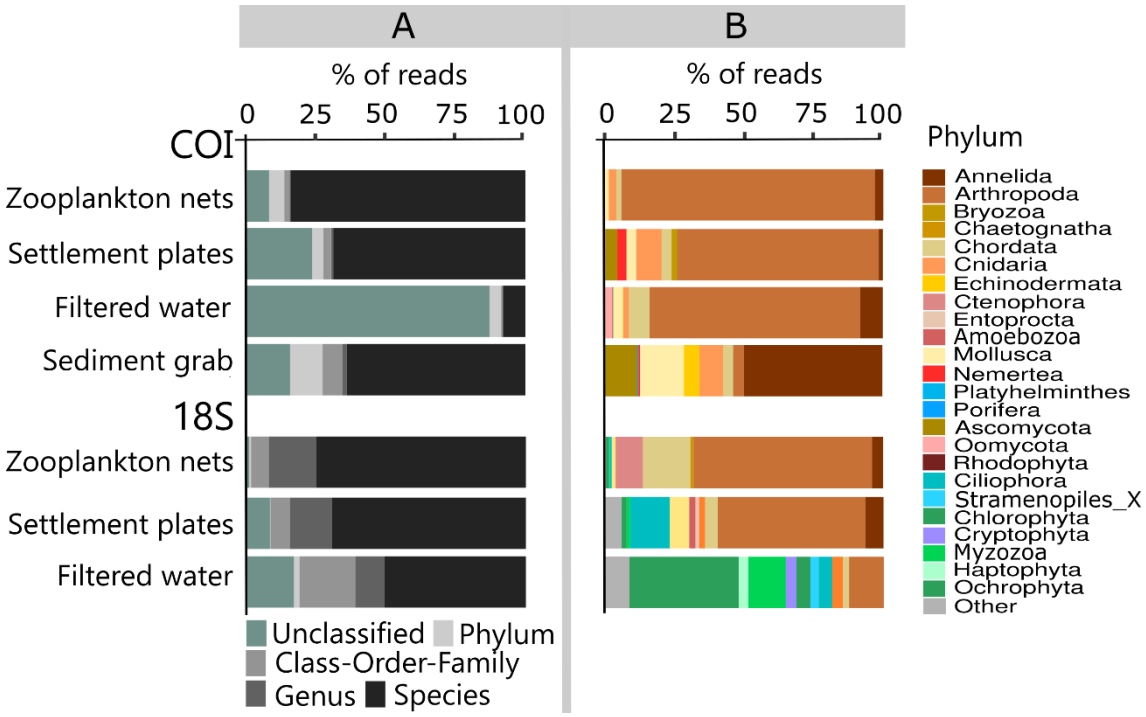


Figure 3. Taxonomic assignment per barcode and per sampling method. A: Percentage of reads assigned to each taxonomic level. B: Relative abundance of reads classified to at least Phylum level for the 10 most abundant phyla.

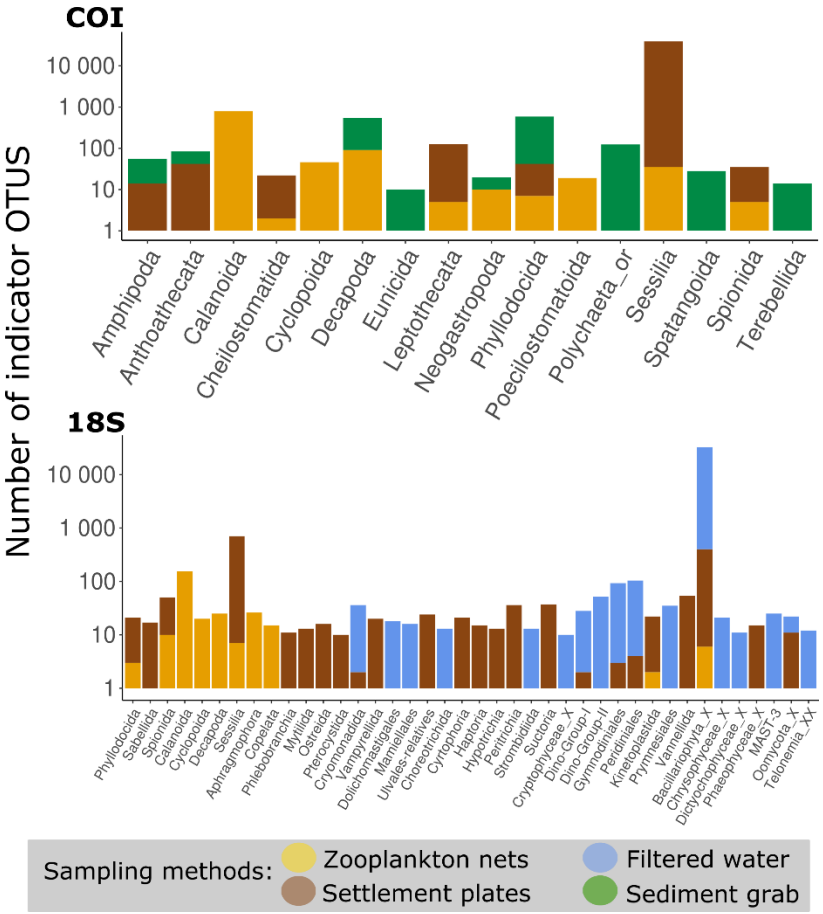


Figure 4. Distribution of indicator OTUs associated to each sampling method at the order level (in log scale) for COI and 18S. Only orders with at least 10 indicator OTUs are shown.

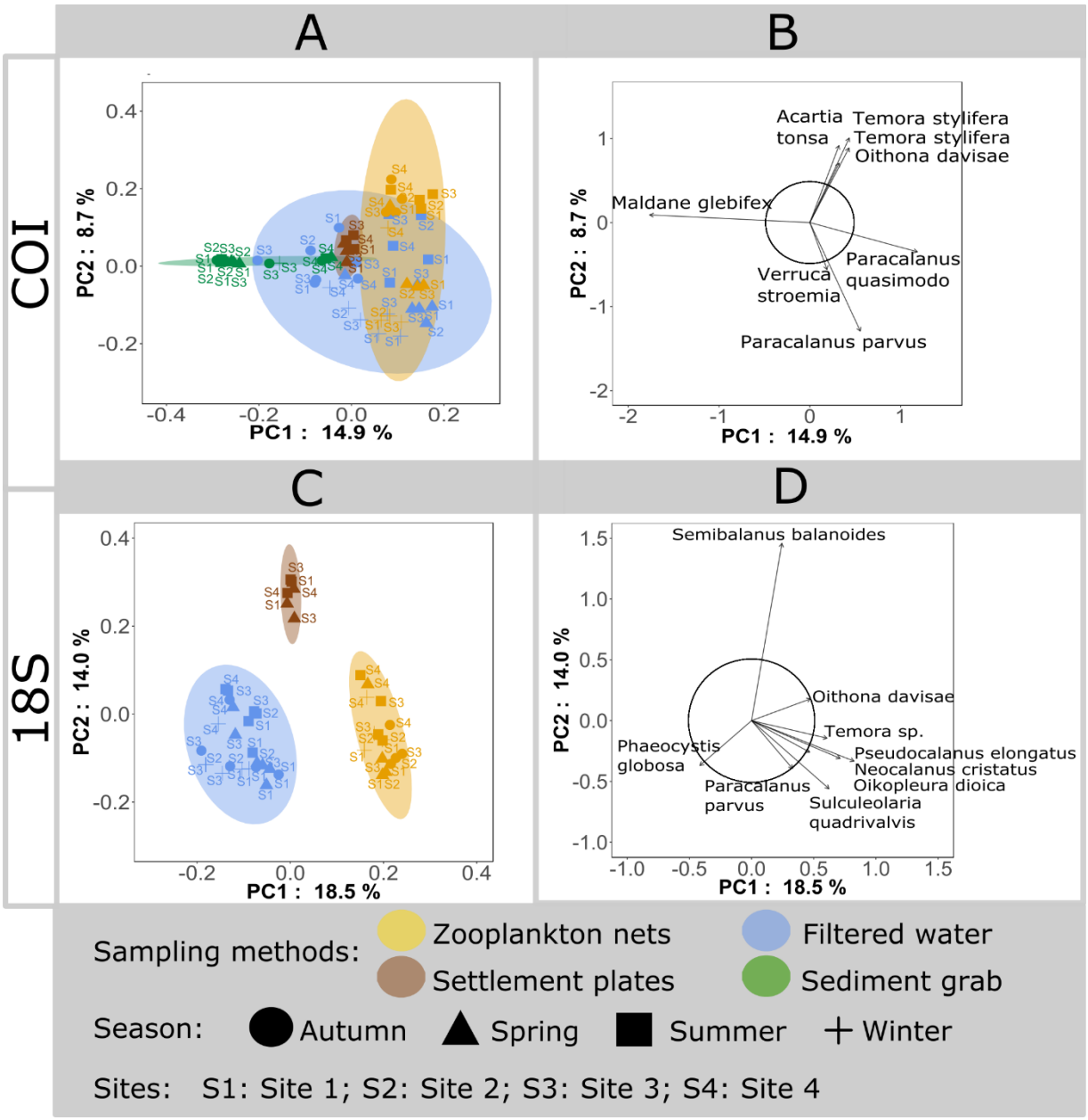


Figure 5. Principal Component Analyses of Hellinger-transformed abundances of OTUs included in the *PBBS targeted taxa dataset* for COI (A-B) and 18S (C-D). A & C: Samples scores in scaling 1 with ellipses representing the 95% confidence dispersion of each sampling method. B & D: OTU scores in scaling 1 with the circle of equilibrium contribution. Only OTUs whose projected length in these two principal component axes exceeding the value of equilibrium contribution are represented.

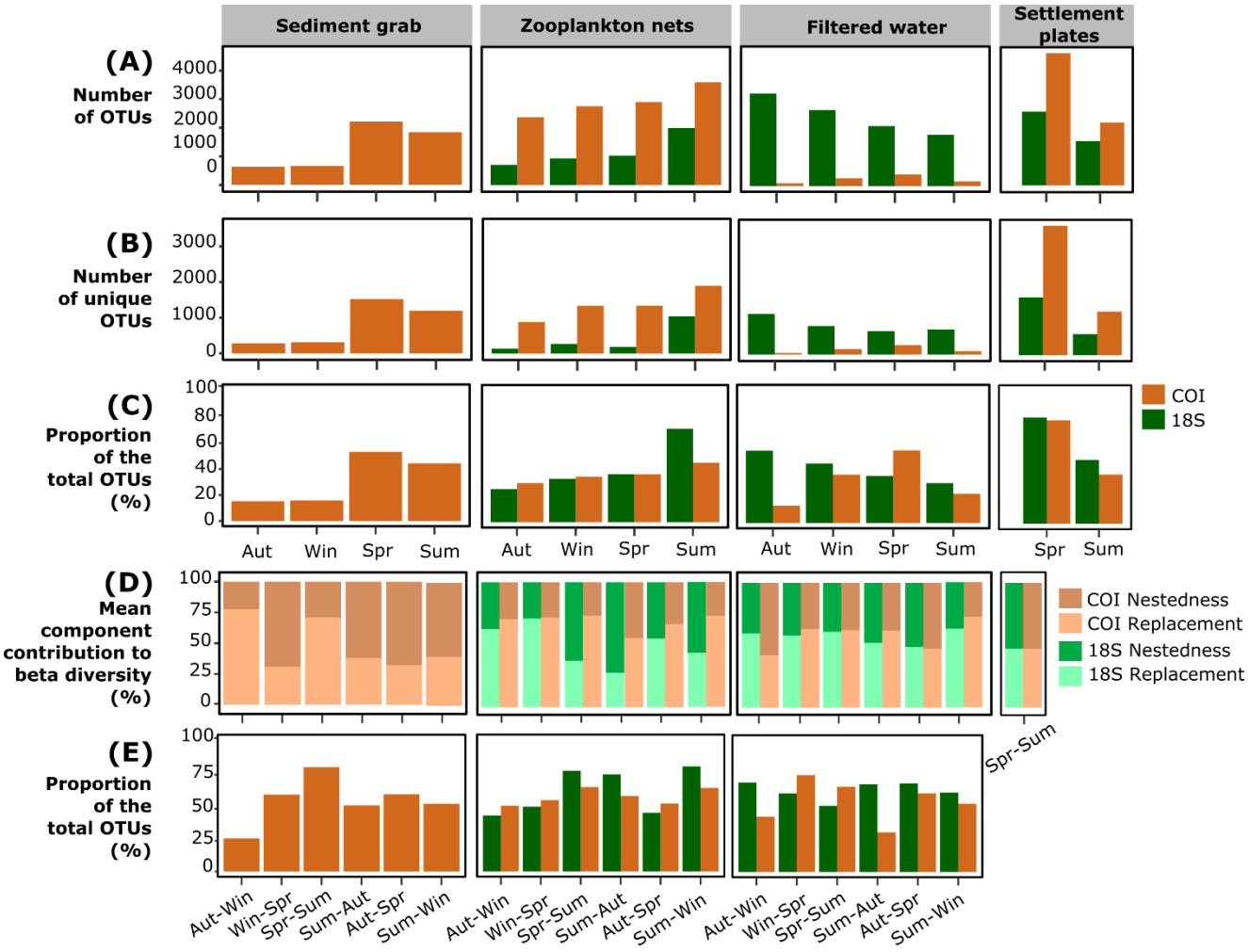


Figure 6. Seasonal variation of alpha and beta diversity for each sampling method with COI and 18S. A: Total OTU richness recovered at each season. B: OTU richness unique to each season. C: Proportion of the total OTU richness detected with one season. D: Decomposition of between-season beta diversity into replacement and nestedness components. E: Proportion of the total OTU richness detected with two seasons.

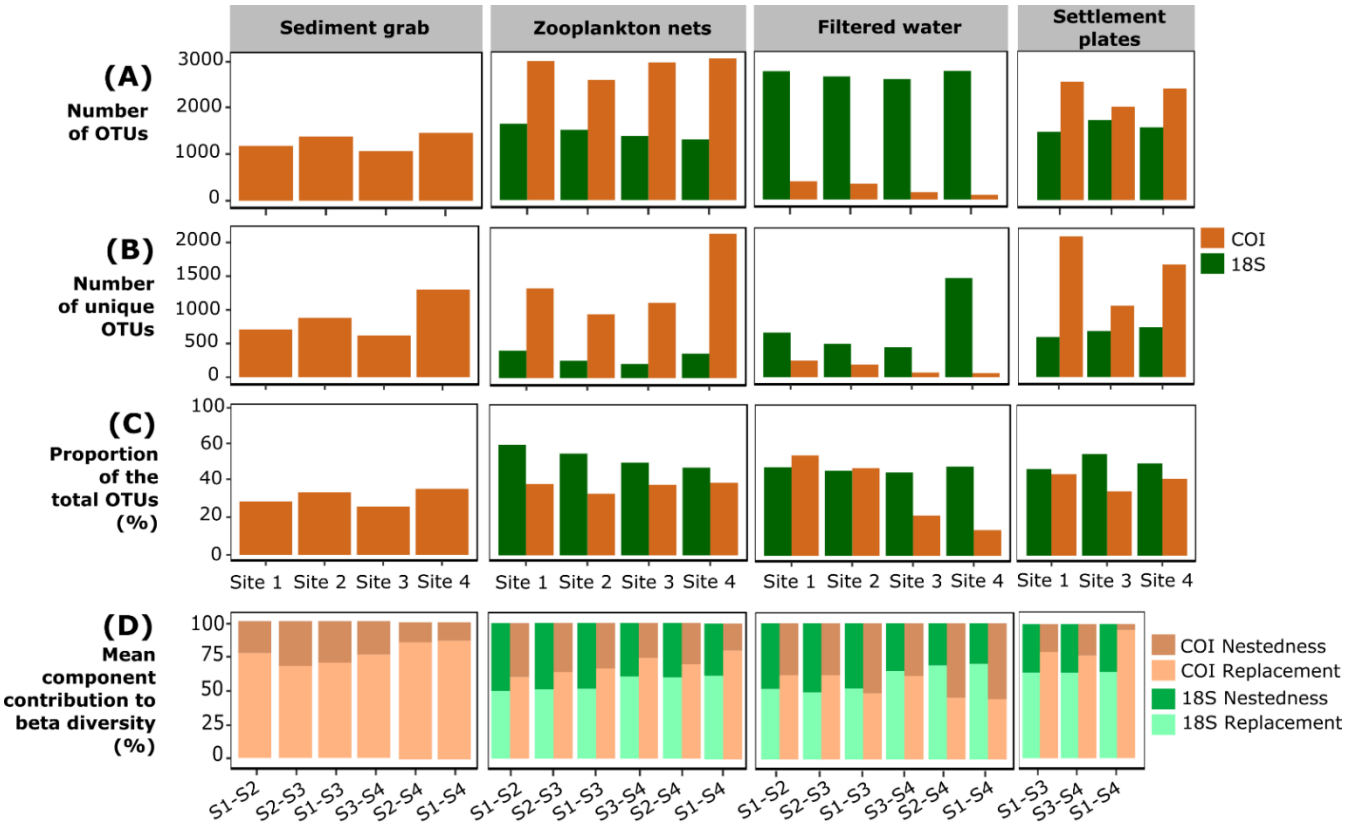


Figure 7. Spatial variation of alpha and beta diversity for each sampling method with COI and 18S. A: Total OTU richness recovered at each site. B: OTU richness unique to each site. C: Proportion of the total OTU richness detected with one site. D: Decomposition of between-site beta diversity into replacement and nestedness components