Considerations for metabarcoding-based port biological baseline surveys aimed at marine non-indigenous species monitoring and risk-assessments

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

Abstract

1. Monitoring introduction and spread of non-indigenous species via maritime transport and performing risk assessments require 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 and the time-consuming morphological approach used for taxonomic identification. Metabarcoding represents a promising alternative for rapid comprehensive port biological baseline surveys, but its application in this context requires further assessments.
2. We applied metabarcoding (based on barcodes of the Cytochrome c oxidase subunit I and of the 18S ribosomal RNA gene) to 192 port samples collected i) from diverse habitats (water column – including environmental DNA and zooplankton, sediment and fouling structures), ii) at different sites (from inner to outer estuary), and iii) during the four seasons of the year.
3. By comparing the biodiversity metrics derived from each sample group, we show that each sampling method resulted in a distinct community profile and that environmental DNA alone cannot substitute for organismal sampling, and that, although sampling at different seasons and locations resulted in higher observed biodiversity, operational results can be obtained by sampling selected locations and seasons.
4. By assessing the taxonomic composition of the samples, we show that metabarcoding data allowed the detection of previously recorded non-indigenous species as well as to reveal presence of new ones, even if in low abundance.

5. *Synthesis and application.* Our comprehensive assessment of metabarcoding for port biological baseline surveys sets the basics for cost-effective, standardized and comprehensive monitoring of non-indigenous species and for performing risk assessments in ports. This development will contribute to the implementation of the recently entered into force 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

Introduction

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

One of these legal frameworks is the International Convention for the Control and Management of Ships' Ballast Water and Sediments (IMO, 2004), which requires that ships treat their ballast water before its release to port, unless they show that the risk of transferring NIS between the donor and recipient ports is limited (David, Gollasch, & Pavliha, 2013). Such risk assessment requires cataloguing biodiversity through Port Biological Baseline Surveys (PBBS), which requires collecting samples using various methods (Kraus et al., 2018) and at different spatial and temporal scales (Lehtiniemi et al., 2015) given the diverse range of habitats (such as soft sediment, water column, or exposed and sheltered artificial structures) allowing presence of distinct organismal groups (such as benthic macrofauna, fouling organisms or planktonic organisms).

Global initiatives have been initiated to standardize sampling for PBBS (HELCOM/OSPAR, 2013; Awad, Haag, Anil, & Abdulla, 2014; Kraus et al., 2018) and are based on morphological taxonomic identification of the species found. Yet, this method lacks accuracy for identification of larvae and eggs, developmental stages at

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

78 DNA metabarcoding (Taberlet, Coissac, Pompanon, Brochmann, & Willerslev,
79 2012), the simultaneous identification of taxa present in a complex environmental
80 sample based on a conserved DNA fragment, is revolutionizing traditional biodiversity
81 monitoring (Creer et al., 2016; Elbrecht, Vamos, Meissner, Aroviita, & Leese, 2017),
82 and could represent the cost-effective and possible to standardize alternative required
83 for PBBS (Lehtiniemi et al., 2015). Metabarcoding can be applied to DNA extracted
84 from bulk organismal samples (community DNA) or to environmental DNA (eDNA),
85 which is extraorganismal DNA released in the environment in form of cells, feces, skin,
86 saliva, mucus, etc. (Shaw, Weyrich, & Cooper, 2016). The later has received particular
87 attention lately for its potential to cost-effectively and noninvasively survey species
88 richness from many ecosystems (Deiner, Fronhofer, Mächler, Walser, & Altermatt,
89 2016), and for being able to detect spatially discrete communities (Jeunen, Knapp,
90 Spencer, Lamare, et al., 2019).

91 Metabarcoding has been extensively explored and validated as an alternative
92 tool to morphological identification (e.g. Aylagas et al. (2016; 2018), and described as a

93 promising tool for NIS monitoring (Comtet, Sandionigi, Viard, & Casiraghi, 2015).
94 Recent studies have highlighted the potential of metabarcoding applied to eDNA to
95 perform port biodiversity surveys (Borrell, Miralles, Huu, Mohammed-Geba, & Garcia-
96 Vazquez, 2017; Grey et al., 2018; Lacoursière-Roussel et al., 2018). Yet, very few
97 studies have performed port surveys based on additional sampling substrates (Brown,
98 Chain, Zhan, MacIsaac, & Cristescu, 2016; Zaiko et al., 2016) and only one has
99 compared biodiversity assessments obtained by the different substrates (Koziol et al.,
100 2019). Additionally, few studies have considered the spatial and temporal variability of
101 ports (*e.g.* eDNA in arctic ports Lacoursière-Roussel et al. (2018) and settlement plates
102 in austral temperate port Zaiko et al. (2016)) but none has fully evaluated the impact
103 of spatial and temporal variability on all the different communities of a port, which is
104 critical to ensure maximum biodiversity recovery in PBBS.

105 Here, we have applied metabarcoding to 192 samples collected from the port
106 of Bilbao (Northern Spain) to evaluate the effect of i) using various sampling methods
107 to capture biodiversity found in different substrates, ii) sampling at different sites from
108 outer to inner estuary, iii) sampling during different seasons of the year, and iv) using
109 alternative genetic markers. Our results provide relevant information for performing
110 non-indigenous species monitoring and risk assessments in ports in response to the
111 International Convention for the Control and Management of Ships' Ballast Water and
112 Sediments.

113

Materials and methods

Sampling

Zooplankton, sediment, fouling and water samples were collected from the port of Bilbao in autumn 2016, winter 2017, spring 2017, and late summer 2017 at four sites (Fig 1 and Appendix 1). Sites 1, 2 and 3 represent the outer parts of the port with busy berths and are characterized by deep waters (10-30 m) and salinity (>30 PSU); site 4 represent s the inner, less busy part of the port and is characterized by low water depth (6-9 m) and lower salinity (>20 PSU). Zooplankton, sediment and fouling sampling was based on the protocol designed by Helcom/Ospar (HELCOM/OSPAR, 2013). Zooplankton samples were collected in vertical tows using Pairovet nets (mesh sizes of 60 and 280 µm) at three points per site, mixing the collected material to have one sample per mesh size per site. The collected zooplankton was grinded with a mortar until no integer organism could be appreciated and was stored in 96% ethanol at -20 °C. Sediment samples (3 samples per site) were collected by sieving material collected from a Van Veen grab (0.07–0.1 m^2) with a 1 mm mesh size sieve. Retained material was processed following Aylagas et al. (2016) based on decantation and homogenization phases, and retrieved benthic macroinvertebrates were stored in 96% ethanol at -20 °C. Homogenization was done by blending the decanted organic material in a PHILIPS hr2095 700 W 2 L glass jar with 96% ethanol until no fragments of animals and others organic materials could be observed. During decantation and homogenization steps, all laboratory equipment used were cleaned between samples by soaking in 10% bleach for 15 min and thoroughly rinsing with deionized water. Fouling samples were collected by placing 15 x 15 cm polyvinyl chloride plates at 1, 3 and 7 m depth in a suspended array in each site. Plates were deployed in winter and

138 spring and recovered in spring and late summer respectively (see Table S1). At
139 recovery, were placed in individual sterile plastic bags, soaked in 96% ethanol and
140 thoroughly scrapped both sides with a scalpel to retrieve fouling organisms attached to
141 it. The detached organisms were then homogenized with a blender (Conair™ Waring™
142 Mini-sample Containers for Blender) in 96% ethanol and stored in 96% ethanol at -20
143 °C. Cleaning procedure was similar as the one performed for sediment samples. At
144 each sampling site, 1 L surface water sample was collected from each sampling point
145 with a bottle. Samples from the same site and from the same sampling method were
146 combined into a single sample by site. In sites 1 and 3, three additional samples were
147 taken at one meter above the bottom using a Niskin bottle and combined. Each 3 L
148 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, 3 L
150 surface water sample from three sampling sites from the ports of Vigo and A Coruña,
151 located in Galicia, Northwestern Spain following the same sampling procedure than for
152 Bilbao by combining 1 L water from three sampling point per sampling site, were
153 collected in March 2017 and filtered as described above.
154
155 *DNA extraction, library preparation and sequencing*
156 Total genomic DNA was isolated from zooplankton samples with the DNeasy blood &
157 tissue kit (QIAGEN), benthic macroinvertebrate and fouling organism samples with,
158 respectively, PowerMax and PowerSoil DNA Isolation Kits (MOBIO) replacing the initial
159 bead-beating step by an overnight incubation at 56°C with proteinase K (0.4 mg/ml).
160 Filter samples were extracted with the DNeasy blood & tissue kit (QIAGEN) following
161 the “SX filters without preservation buffer“ developed by Spens et al. (2017). Negative

162 controls (preservation buffer) were included in each batch of DNA extraction and
163 followed the same procedure as all other samples. DNA concentration was measured
164 with the Quant-iT dsDNA HS assay kit (Thermo Scientific) using a Qubit 2.0 Fluorometer
165 (Life Technologies), purity was inferred from 260/280 and 260/230 absorbance ratios
166 with the ND-1000 Nanodrop (Thermo Scientific), and integrity was assessed by
167 electrophoresis in 0.7% agarose. Two primer pairs widely used to assess eukaryotic
168 biodiversity (Hebert, Ratnasingham, & de Waard, 2003; Amaral-Zettler, McCliment,
169 Ducklow, & Huse, 2009) were used: mlCOIintF/dgHCO2198 (COI primers), targeting a
170 313 bp fragment of the Cytochrome c oxidase subunit I (COI) gene (Meyer, 2003; Leray
171 et al., 2013) and 1389F/1510R (18S primers), targeting a variable length fragment (87
172 to 186 bp) of the V9 hypervariable region of the 18S rRNA gene (Amaral-Zettler et al.,
173 2009). DNA extracted from zooplankton, fouling organism and filter samples were
174 amplified with both primer pairs, whereas DNA from sediment was only amplified with
175 the COI primers. PCR amplifications were performed in two rounds. For the first PCR, 1
176 µl of genomic DNA (5 ng/µl) was added to a mix consisting in 5 µl of 2X Phusion Master
177 Mix (Thermo Scientific), 0.2 µl of each primer (0.2 µM) and 2.6 µl of MilliQ water. For
178 the 18S primers, PCR conditions consisted on an initial 3 min denaturation step at 98
179 °C; followed by 25 cycles of 10 s at 98 °C, 30 s at 57 °C and 30 s at 72 °C and finally 10
180 min at 72 °C. For the COI primers, PCR conditions consisted of an initial 3 min
181 denaturation step at 98 °C; followed by 35 cycles of 10 s at 98 °C, 30 s at 46 °C and 45 s
182 at 72 °C and finally 5 min at 72 °C. Negative controls of PCR (no template) were
183 included within each set of PCRs. For each DNA extract, three PCR amplifications were
184 performed and pooled. Once purified using AMPure XP beads (Beckman Coulter), the
185 mixed PCR products were used as template for the generation of dual-indexed

186 amplicons in a second PCR round following the "16S Metagenomic Sequence Library
187 Preparation" protocol (Illumina) using the Nextera XT Index Kit (Illumina). Multiplexed
188 PCR products were purified again using the AMPure XP beads, quantified using Quant-
189 iT dsDNA HS assay kit and a Qubit 2.0 Fluorometer (Life Technologies), normalized to
190 equal concentration and sequenced using the 2 x 300 paired-end MiSeq (Illumina).
191 Reads were demultiplexed based on their barcode sequences.
192
193 *Raw read preprocessing, clustering and taxonomic assignment*
194 After quality checking of demultiplexed paired-end reads with FASTQC (Andrews,
195 2010), forward and reverse primers were removed using Cutadapt (Martin, 2011) with
196 the anchored 5' adapter and for "paired end reads" options and with the linked
197 adapter option for COI and 18S respectively. Forward and reverse reads were merged
198 using PEAR (Zhang, Kobert, Flouri, & Stamatakis, 2014) with a minimum sequence
199 overlap of 217 bp and maximum amplicon length of 313 bp and of 80 and 190 bp for
200 COI and 18S barcodes respectively. Merged reads with average Phred quality score
201 lower than 20 were removed with Trimmomatic (Bolger, Lohse, & Usadel, 2014). Using
202 Mothur (Schloss et al., 2009), sequences without ambiguous bases were aligned to
203 BOLD (https://www.boldsystems.org) or SILVA (https://www.arb-
204 silva.de/documentation/release-132/) for COI and 18S respectively, and only those
205 covering the barcode region were kept. Chimeras, detected using *de novo* mode of
206 UCHIME (Edgar, Haas, Clemente, Quince, & Knight, 2011), were removed, and
207 remaining reads were clustered into OTUs using Swarm 2.2.1 with the step-by-step
208 aggregation clustering algorithm implemented with default settings (Mahé, Rognes,
209 Quince, de Vargas, & Dunthorn, 2014). SWARM algorithm does not rely on a fixed

210 threshold for delimiting OTUs which is pertinent when performing PBBS where highly
211 diverse biodiversity can be found. “Singleton” OTUs, composed by a single read, were
212 removed. No rarefaction was performed to avoid decreasing sensitivity by choosing an
213 arbitrary minimum library size (McMurdie and Holmes, 2014). The remaining OTUs
214 were taxonomically assigned using the Naïve Bayes Classifier (Wang, Garrity, Tiedje, &
215 Cole, 2007) using BOLD (accessed in May 2018) or PR2 (release 4.10.0) databases for
216 COI and 18S respectively.
217
218 *Community analyses*
219 Apart from the complete dataset, we created two subsets: a taxa targeted through
220 PBBS dataset, including only reads classified to the class level and excluding those
221 matching to non-targeted groups for PBBS such as Mammalia, Aves, Insecta,
222 Collembola, Arachnida and all classes of Fungi; and a NICS dataset, including only reads
223 matching either the 68 Non-Indigenous and Cryptogenic Species (NICS) previously
224 detected in the port of Bilbao (Martínez & Adarraga, 2006; Adarraga & Martínez, 2011,
225 2012; Butrón, Orive, & Madariaga, 2011; Zorita et al., 2013; Tajadura, Bustamante, &
226 Salinas, 2016) or the 1083 species present in the AquaNIS database (AquaNIS. Editorial
227 board., 2015). Most analyses were conducted using RStudio (Team, 2015) with *vegan*
228 (Oksanen et al., 2013), *adespatial* (Dray et al., 2018) and *indicspecies* (De Cáceres &
229 Legendre, 2009) libraries. Indicator species analyses (Dufrene & Legendre, 1997) were
230 performed on 1) the taxa targeted through PBBS dataset to identify indicator taxa of
231 each sampling method and 2) the NICS dataset to identify non-indigenous indicator
232 taxa of each sampling site. These analyses were based on the IndVal index calculated
233 as the product of the degree of specificity (measuring the uniqueness to a sampling

234 method or site) and the degree of fidelity (measuring the frequency of occurrence
235 within a sampling method or site) of an OTU to a given sampling condition. Statistical
236 significance of associations was assessed by performing 10,000 permutations. The
237 effects of season and locality on taxa targeted through PBBS communities were tested
238 for significance using a permutational multivariate analysis of variance (PERMANOVA)
239 after checking for multivariate homogeneity of group dispersions (betadisper).
240 PERMANOVA and betadisper were performed on Euclidean distances on Hellinger-
241 transformed OTU abundances (Hellinger distances), which are appropriate for
242 community ordination and clustering (Legendre & Gallagher, 2001) and on Jaccard
243 dissimilarities based on OTU presence/absence. The contribution of replacement
244 (changes in OTU identity) and nestedness (richness differences where one sample is a
245 subset of a richer sample) to beta diversity of taxa targeted through PBBS between
246 seasons and between sites was computed using the relativized nestedness index of
247 Podani and Schmera (2011) based on Jaccard dissimilarities matrix. For each sampling
248 method, replacement and nestedness were calculated between pairwise comparisons
249 of 1) samples belonging to the same site but sampled at different seasons (season
250 variation of beta diversity), and 2) samples belonging to the same season but sampled
251 at different sites (spatial variation of beta diversity). The mean proportion of
252 nestedness and replacement contribution to beta-diversity between sites and between
253 seasons was then calculated. For all OTUs assigned to a Non-Indigenous and
254 Cryptogenic Species, we blasted their representative sequences against BOLD and PR2
255 databases respectively for COI and 18S barcodes.
256

257 **Results**

258 *Communities retrieved by each sampling method and genetic marker*

259 A total of 5,718,639 and 7,055,675 COI and 18S barcode reads were kept for
260 analysis (Table S1). Negative controls showed little contamination (with only 164, 66
261 and 213 of COI and 2031 and 163 18S quality filtered reads). The reads corresponding
262 to the 192 samples collected at the port of Bilbao resulted in 40,318 and 20,473 OTUs
263 for COI and 18S respectively. For all sampling methods, OTU accumulation curves of
264 OTUs against reads approached saturation, suggesting that adding more sequencing
265 effort would provide limited increase in diversity (Figs 2A and Appendix 2). The
266 majority of the OTUs, 89% for COI and 73% for 18S, were unique to one sampling
267 method (Fig. 2B), and, for both barcodes, but particularly for 18S, filtered water (aimed
268 at retrieving eDNA from macroorganisms and microbial eukaryotes) resulted in higher
269 OTU richness and unique OTUs than other sampling methods (Figs 2A, B and Appendix
270 2). Different sampling methods retrieved distinct biological communities as observed
271 with the PCA (Fig. 2C) and confirmed with PERMANOVA analyses (COI: R^2=0.28, p-
272 value= 9.99e-05; 18S: R^2=0.30, p-value=9.99e-05).

273 The percentage of reads assigned to species or higher taxonomic levels differed
274 between sampling methods and barcodes. The majority of the reads were assigned to
275 species level for zooplankton nets (COI: 84% and 18S: 75%), settlement plates (COI:
276 69% and 18S: 69%) and sediment grabs (COI: 64%); for filtered water, although 51% of
277 reads were assigned to species level with 18S, only 8% could be assigned to a species
278 with COI (Fig. 3A). Filtered water analyzed with COI had also the largest proportion of
279 reads that could not be assigned to phylum (87%). This was due to the non-specific

280 amplification of prokaryotic and non-target eukaryotic DNAs (Appendix 3), which is
281 known for COI (Collins et al., 2019).

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

288

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

290 A total of 16,828 and 9,091 OTUs were kept as taxa targeted through PBBS for
291 COI and 18S respectively. From those, indicator analysis identified respectively 2,600
292 (15%) and 1,700 (19%) OTUs significantly associated to one of the sampling methods.
293 Settlement plates were associated with 1,268 (COI) and 808 (18S) OTUs, zooplankton
294 nets with 1,001 (COI) and 315 (18S) OTUs, and sediment grabs with 238 (COI) OTUs.
295 Only 12 OTUs (9 of which were metazoans) were indicators of filtered water with COI
296 while 580 (3 of which were metazoans), were associated to this sampling method with
297 18S. We observed expected strong associations of some taxa to one particular
298 sampling method: fouling bivalves (Ostreoida, Mytilida) with settlement plates,
299 copepods (Calanoida, Cyclopoida) with zooplankton nets and sea urchins (Spatangoida)
300 with sediment grabs (Fig. 4). Yet, less obvious associations were also observed such as
301 barnacles (Sessilia) and two polychaeta orders (Spionida, Phyllodocida) with
302 zooplankton nets or dinoflagellates (Peridiniales and Gymnodiniales) with settlement

303 plates, illustrating the advantages of a complementary sampling approach to recover
304 the diversity of these taxonomic groups.

305 The need for complementarity of sampling methods was further confirmed by
306 PCA on PBBS target taxa (Fig. 5). For both barcodes, expected taxonomic differences
307 were found between the sampling methods: sediment grabs were characterized by the
308 polychaeta *Maldane glebifex* (Fig. 5B), zooplankton nets were distinguished by several
309 copepods (Fig. 5B, D) and settlement plates, by the encrusting *Semibalanus balanoides*
310 (Fig. 5D). Filtered water communities were characterized differently according to each
311 barcode. For COI, filtered water samples were not differentiated as a distinct group
312 and were in general close to those retrieved with zooplankton samples collected at the
313 same season (Fig. 5A). In contrast, for 18S, filtered water was different due to the
314 presence of the phytoplankton species *Phaeocystis globosa* (Fig. 5D); yet, when
315 targeting only metazoan taxa, the patterns observed with COI were similarly to the
316 ones retrieved with 18S (Appendix 4). For filtered samples, the proportion of metazoan
317 OTUs detected was largely different compared to the other methods (Appendix 5). For
318 instance, Decapoda orders had a low diversity with filtered water (COI: 0,7%; 18S: 7%),
319 whereas more diversity could be detected for Calanoida (COI: 8%; 18S: 35%), Sabellida
320 (COI: 15%; 18S: 28%) and Leptothecata (COI: 13%; 18S: 58%). In general, filtered water
321 recovered the smallest metazoan diversity.

322

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

324 Seasonal variations contributed significantly to differences in community
325 composition of zooplankton nets and filtered water but not of sediment grab (Table
326 S2). Total richness and unique richness varied between seasons for all sampling

327 methods (Fig. 6A, B). Spring and late summer were generally richer than autumn and
328 winter, excepted for filtered water communities represented with 18S. For sediment
329 grabs and zooplankton nets with 18S, a strong proportion of taxa found in autumn and
330 winter were a subset of taxa retrieved in late summer and spring. Indeed, the
331 nestedness component of sediment grab assemblages represented between 60 to 70%
332 of total compositional variations among these pairs of seasons (Fig. 6D). For
333 zooplankton net assemblages, 63 and 72% of nestedness was observed between late
334 summer and spring and between late summer and autumn respectively, suggesting
335 that late summer recovered a majority of spring and autumn diversity (Fig. 6D). Thus,
336 the combination of spring and late summer for sediment grabs retrieved 85% of the
337 total OTU richness over the four seasons, while for zooplankton nets, in late summer
338 alone 71% was retrieved (Fig. 6C, E). For filtered water and settlement plates, seasonal
339 community changes were driven by both OTU replacement and nestedness, with
340 similar relative contributions (Fig. 6D). The seasonal influence observed in intra-port
341 samples was also observed between ports (Appendix 6). The differences in
342 communities between Bilbao, A Coruña and Vigo (ports belonging to the same
343 ecoregion but separated by over 500 km) during the same season was generally
344 smaller than that between seasons in the same port, indicating that communities were
345 driven by seasonality rather than location (Appendix 6 and Table S3). This pattern was
346 more pronounced with 18S than with COI.

347 Locality within the Bilbao port appeared to impact on benthic assemblages
348 (Table S2), since sites outside the estuary (sites 1 to 3) were different from site 4 inside
349 the estuary; mainly, the polychaete *Maldane glebifex* was less abundant in site 4 (Fig.
350 5B). Regarding zooplankton nets, site 4 was the main driver of difference in

351 communities as, when not considered, no significant differences between sites were
352 observed (Table S2). Each site harbored a similar proportion of the total OTU richness
353 found by each sampling method (Fig. 7C). This proportion did not exceed 60%. Indeed,
354 OTU replacement contributed more to community variation among sites than
355 nestedness, especially with sediment grab and settlement plates (Fig. 7D). OTU
356 replacement was more important when comparing sites from outside the estuary
357 (sites 1, 2 and S) with site 4 inside the estuary, while it contributed less to community
358 variation among sites 1,2 and 3. This is congruent with site 4 having generally more
359 unique OTUs than the other sites (Fig. 7B). An exception was observed for filtered
360 water with COI, where site 4 had the lowest OTU richness and unique OTU richness,
361 and where nestedness contributed more to the variation in community composition in
362 comparison to sites 1 and 2 (Fig. 7A, B, D).
363
364 *Detection of Non-Indigenous and Cryptogenic Species (NICS)*
365 Our port baseline biological survey detected 79 putative NICS, among which 29
366 of the 68 previously recorded in the port Bilbao were found (Tables S4, 5 and 6). Most
367 of the other species (43) were previously detected NICS inside the port's Large Marine
368 Ecoregion (LME) but, for 7 of them (*Ammothea hilgendorfi, Bugulina fulva,*
369 *Grandidierella japonica, Melita nitida, Neodexiospira brasiliensis, Pseudochattonella*
370 *verruculosa, Tubificoides pseudogaster*), it was, to our knowledge, the first report in
371 the LME. The indicator species analysis performed on reads corresponding to non-
372 indigenous species revealed that site 4 was associated to the highest number of non-
373 indigenous and cryptogenic species, which were *Oithona davisae*, *Acartia tonsa* and
374 *Allita succinea*, found by both barcodes, *Ficopomatus enigmaticus,* found with 18S and

375 *Amphibalanus eburneus*, *Grandidierella japonica*, *Polydora cornuta*, *Austrominius*
376 *modestus*, *Monocorophium acherusicum*, *Xenostrobus securis,* found with COI.
377 Regarding the three other sites, for COI, site 1 was associated to *Clytia hemisphaerica*
378 and to *Balanus trigonus*, site 2, to *Clytia hemisphaerica* and *Mytilus edulis* and site 3, to
379 *Balanus trigonus*. For 18S, only site 1 was associated to *Gymnodinium aureolum.*
380

381 **Discussion**

382 *Metabarcoding-based port baseline surveys require a combination of sampling*
383 *methods and should not rely solely on eDNA*

384 Most of the past attempts on using metabarcoding for port monitoring have
385 relied on a single 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 using different sampling methods (Koziol et al., 2019). In agreement
388 with the latter, our metabarcoding analysis shows that each sampling method
389 recovered a distinct subset of the port community, and that, despite some taxa being
390 expectedly associated to a given sampling method, the total diversity, even within
391 specific taxonomic groups, was only recovered by combining different methods.
392 Interestingly, despite the documented potential of eDNA metabarcoding to capture a
393 large fraction of the macroorganismal diversity with limited effort (Bista et al., 2017;
394 O'Donnell et al., 2017), including detection of non-indigenous taxa (Klymus, Marshall,
395 & Stepien, 2017) and port surveys (Borrell et al., 2017; Grey et al., 2018; Lacoursière-
396 Roussel et al., 2018), our analyses show that, compared to bulk sample metabarcoding,
397 eDNA metabarcoding recovered only a subset of the metazoan diversity and did not
398 provide additional information on targeted groups. This confirms previous findings in

399 coral reef sites (Djurhuus et al., 2018) and freshwater streams (Macher et al., 2018)
400 and recently in ports (Koziol et al., 2019), suggesting that despite requiring less
401 sampling effort, COI and 18S based metabarcoding of eDNA obtained from filtered
402 water should not be used to replace the need of obtaining bulk samples (using
403 zooplankton nets, grabs and settlement plates) for PBBS taxa detection. Increasing
404 evidences showed that the use of universal and degenerated COI primers is leading to
405 non-specific amplification of prokaryotic and non-target eukaryotic DNA (Collins et al.,
406 2019). Yet, in this context, at least two alternatives are possible to improve eDNA-
407 based biodiversity assessments: i) increasing sequencing depth (Grey et al., 2018) or ii)
408 using group specific markers (Jeunen, Knapp, Spencer, Taylor, et al., 2019). Yet, none
409 of them ensures full biodiversity recovery and both significantly increase costs. Thus,
410 the decision between increasing sampling depth and/or using multiple group specific
411 primers for eDNA and adding multiple substrate sampling needs to be carefully
412 considered and will be an important area of future research to optimize
413 metabarcoding-based port monitoring.
414
415 *Metabarcoding-based port baseline surveys should include spatiotemporal sampling*
416 While it is expected that increased temporal coverage will retrieve more taxa,
417 the HELCOM/OSPAR protocol, a widely applied protocol developed for PBBS
418 (http://jointbwmexemptions.org/ballast_water_RA/apex/f?p=104:13), limits sampling
419 to late summer for sediment and fouling, and to spring and late summer for plankton
420 in order to reduce costs. Yet, so far, no studies using metabarcoding have been
421 performed to support this decision. Here we show that for sediment and fouling,
422 spring sampling provides higher diversity than late summer. Although sampling in late

423 summer could be appropriate for morphological taxonomy because of the more
424 abundance of adult individuals, for metabarcoding, sampling in spring is preferable
425 because during this season i) sizes of organisms are less variable and thus
426 metabarcoding is less likely to under detect small organisms (Elbrecht, Peinert, &
427 Leese, 2017) and ii) species diversity is at its maxima due to being a high recruitment
428 period with abundant organisms at early life stages (Bijleveld et al., 2018). For
429 zooplankton diversity, our results show that the HELCOM/OSPAR sampling in spring
430 and late summer produces the highest diversity. However, the data obtained from
431 filtered water were inconclusive. Previous studies have already shown that seasonal
432 variations are important considerations for eDNA studies (Lacoursière-Roussel et al.,
433 2018), but further characterizations are needed over multiple years to design an
434 adequate protocol for maximizing biodiversity recovery with eDNA.

435 Concerning spatial sampling, for all sampling methods, OTU replacement
436 generally contributed more to community variation among sites than nestedness,
437 suggesting that spatially comprehensive sampling is crucial to recover the port´s
438 biodiversity. Interestingly, site 4 was not only the most different from all four sites but
439 was also the one that recorded the largest number of NIS. In coherence with our
440 findings, it has been observed that brackish environments favors NIS settlement (Zorita
441 et al., 2013) because these species usually support wider range of salinity (Cardeccia et
442 al., 2018). Thus, samples for port monitoring should include those with a wide range of
443 abiotic conditions and covering priority sampling, such as highly active ship berths,
444 potential reservoirs of newly arrived NIS (Hewitt & Martin, 2001).

445

446 *Metabarcoding provides valuable information on Non-Indigenous and Cryptogenic*
447 *Species (NICS)*
448 Our metabarcoding port biological baseline survey detected NICS previously recorded
449 in the port of Bilbao and NICS known to be present in the port´s marine ecoregion.
450 Importantly, it also unveiled presence of seven NICS for which no records exist so far in
451 this ecoregion, highlighting the potential of metabarcoding for early NICS detection.
452 Nonetheless, not all NICS species previously recorded in Bilbao were found by our
453 analyses. This might be due to these species not being present in the port at the time
454 of our survey, to biases of the metabarcoding process such as differential DNA
455 extraction, primer non-specificity, etc (Xiong, Li, & Zhan, 2016), or to database
456 incompleteness. Primer bias and/or database incompleteness could also explain why
457 35 and 42% of NICS were uniquely detected with COI and 18S respectively. Yet, from
458 the total species included in the AquaNIS database (n=1083), only 460 and 369 are
459 included in the BOLD and PR2 databases respectively. This stresses the need of
460 increasing reference databases and of using multiple universal primers for species
461 detection (Grey et al., 2018). Importantly, our study confirmed that metabarcoding can
462 detect species occurring at low abundance (Pochon, Bott, Smith, & Wood, 2013) as we
463 found the non-indigenous amphipod *Melita nitida* with only 18 reads while it took
464 intensive surveys in 2013, 2014 and 2016 in 3 distinct sampling regions of the Bay of
465 Biscay to record 76 individuals (Gouillieux, Lavesque, Blanchet, & Bachelet, 2016).
466

467 **Conclusions**

468 The Ballast Water Management Convention, entered into force in September 2017,
469 aims at preventing the spread of non-indigenous species from one region to another
470 (IMO, 2004). Yet, the implementation of this convention still requires technological
471 developments for assessing compliance and granting exemptions, for which genetic
472 methods have been suggested promising (Rey, Basurko, & Rodríguez-Ezpeleta, 2018).
473 Based on the comparative analysis of 192 samples assessing the use of alternative
474 sampling methods, of sampling at different seasons and at different port locations, we
475 have demonstrated the suitability of metabarcoding for port biodiversity surveys and
476 NIS monitoring and settled the guidelines for future studies. We show that i)
477 combining two pairs of universal primers provides a more holistic view of the port
478 biodiversity, ii) a combination of sampling methods is necessary to recover the
479 different taxonomic groups, iii) environmental DNA cannot replace traditional
480 sampling; vi) sampling should take place in spring and late summer preferably and v)
481 spatial coverage should cover the port's salinity gradient. Considering the cost-
482 effectiveness of metabarcoding with respect to morphological identification, these
483 guidelines and considerations are particularly relevant for performing the risk
484 assessment required for granting exemptions within the International Convention for
485 the Control and Management of Ships' Ballast Water and Sediments.

486

487

488 **Acknowledgments**

489 This manuscript is a result of the Aquainvad-ED (AQUAtic INVADers: Early Detection,
490 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.

Authors' contributions

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

Data Accessibility

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

512 **References**

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

516 Adarraga, I., & Martínez, J. (2012). First record of the invasive brackish water mytilid
517 *Limnoperna securis* (Lamarck, 1819) in the Bay of Biscay. *Aquatic Invasions*,
518 *7*(2).

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

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

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

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

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

Aylagas, E., Borja, Á., Irigoien, X. & Rodríguez-Ezpeleta, N. (2016b). Benchmarking DNA Metabarcoding for biodiversity-Based monitoring and Assessment. *Frontiers in Marine Science*, *3*. doi:10.3389/fmars.2016.00096

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 (1),* 194-202

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

Bista, I., Carvalho, G. R., Walsh, K., Seymour, M., Hajibabaei, M., Lallias, D., … Creer, S. (2017). Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. *Nature Communications*, *8*, 14087. doi:10.1038/ncomms14087

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

Borrell, Y. J., Miralles, L., Huu, H. D., Mohammed-Geba, K., & Garcia-Vazquez, E. (2017). DNA in a bottle—Rapid metabarcoding survey for early alerts of invasive species in ports. *PLOS ONE*, *12*(9), e0183347. doi:10.1371/journal.pone.0183347

Bott, N. J. (2015). Detection of Invasive Species. In S.-K. K. Prof (Ed.), *Springer Handbook of Marine Biotechnology* (pp. 441–453). Springer Berlin Heidelberg. doi:10.1007/978-3-642-53971-8_16

Brown, E. A., Chain, F. J. J., Zhan, A., MacIsaac, H. J., & Cristescu, M. E. (2016). Early detection of aquatic invaders using metabarcoding reveals a high number of non-indigenous species in Canadian ports. *Diversity and Distributions*, *22*(10), 1045–1059. doi:10.1111/ddi.12465

Butrón, A., Orive, E., & Madariaga, I. (2011). Potential risk of harmful algae transport by ballast waters: The case of Bilbao Harbour. *Marine Pollution Bulletin*, *62*(4), 747–757.

Cardeccia, A., Marchini, A., Occhipinti-Ambrogi, A., Galil, B., Gollasch, S., Minchin, D., … Ojaveer, H. (2018). Assessing biological invasions in European Seas: Biological traits of the most widespread non-indigenous species. *Estuarine, Coastal and 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.

Collins, R. A., Bakker, J., Wangensteen, O. S., Soto, A. Z., Corrigan, L., Sims, D. W., … Mariani, S. (2019). Non-specific amplification compromises environmental DNA

580 metabarcoding with COI. *Methods in Ecology and Evolution*, *10*(11), 1985–2001.
581 doi:10.1111/2041-210X.13276
582 Comtet, T., Sandionigi, A., Viard, F., & Casiraghi, M. (2015). DNA (meta) barcoding of
583 biological invasions: a powerful tool to elucidate invasion processes and help
584 managing aliens. *Biological Invasions*, *17*(3), 905–922.
585 Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W. K., … Bik, H. M.
586 (2016). The ecologist's field guide to sequence-based identification of
587 biodiversity. *Methods in Ecology and Evolution*, *7*(9), 1008–1018.
588 doi:10.1111/2041-210X.12574
589 David, M., Gollasch, S., & Pavliha, M. (2013). Global ballast water management and the
590 "same location" concept: a clear term or a clear issue? *Ecological Applications*,
591 *23*(2), 331–338. doi:10.1890/12-0992.1
592 De Cáceres, M. D., & Legendre, P. (2009). Associations between species and groups of
593 sites: indices and statistical inference. *Ecology*, *90*(12), 3566–3574.
594 Deiner, K., Fronhofer, E. A., Mächler, E., Walser, J.-C., & Altermatt, F. (2016).
595 Environmental DNA reveals that rivers are conveyer belts of biodiversity
596 information. *Nature Communications*, *7*, 12544. doi:10.1038/ncomms12544
597 Djurhuus, A., Pitz, K., Sawaya, N. A., Rojas-Márquez, J., Michaud, B., Montes, E., …
598 Breitbart, M. (2018). Evaluation of marine zooplankton community structure
599 through environmental DNA metabarcoding: Metabarcoding zooplankton from
600 eDNA. *Limnology and Oceanography: Methods*, *16*(4), 209–221.
601 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

environmental DNA metabarcoding surveys. *Scientific Reports*, *8*(1). doi:10.1038/s41598-018-27048-2

Hebert, P. D., Ratnasingham, S., & de Waard, J. R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London B: Biological Sciences*, *270*(Suppl 1), S96–S99.

HELCOM/OSPAR. (2013). Joint HELCOM/OSPAR Guidelines for the Contracting Parties of OSPAR and HELCOM on the granting of exemptions under International Convention for the Control and Management of Ships' Ballast Water and Sediments, Regulation A-4 (OSPAR Agreement 2013-09), 45 pp.

Hewitt, C. L., & Martin, R. B. (2001). Revised protocols for baseline port surveys for introduced marine species–design considerations, sampling protocols and taxonomic sufficiency. *Centre for Research on Introduced Marine Pests Technical Report*, (22).

IMO. (2004). International Convention for the Control and Management of Ships' Ballast Water and Sediments, 2004. International Maritime Organization, 13 Feb 2004, London.

Jeunen, G.-J., Knapp, M., Spencer, H. G., Lamare, M. D., Taylor, H. R., Stat, M., … Gemmell, N. J. (2019). Environmental DNA (eDNA) metabarcoding reveals strong discrimination among diverse marine habitats connected by water movement. *Molecular Ecology Resources*, *19*(2), 426–438. doi:10.1111/1755-0998.12982

Jeunen, G.-J., Knapp, M., Spencer, H. G., Taylor, H. R., Lamare, M. D., Stat, M., … Gemmell, N. J. (2019). Species-level biodiversity assessment using marine

environmental DNA metabarcoding requires protocol optimization and standardization. *Ecology and Evolution*, *9*(3), 1323–1335. doi:10.1002/ece3.4843

Klymus, K. E., Marshall, N. T., & Stepien, C. A. (2017). Environmental DNA (eDNA) metabarcoding assays to detect invasive invertebrate species in the Great Lakes. *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

Legendre, P. & Gallagher, E. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia, 129 (2),* 271-280

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

673 Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., … Machida, R. J.
674 (2013). A new versatile primer set targeting a short fragment of the
675 mitochondrial COI region for metabarcoding metazoan diversity: application for
676 characterizing coral reef fish gut contents. *Frontiers in Zoology*, *10*, 34.
677 doi:10.1186/1742-9994-10-34
678 Macher, J.-N., Vivancos, A., Piggott, J. J., Centeno, F. C., Matthaei, C. D., & Leese, F.
679 (2018). Comparison of environmental DNA and bulk-sample metabarcoding
680 using highly degenerate cytochrome c oxidase I primers. *Molecular Ecology*
681 *Resources*. doi:10.1111/1755-0998.12940
682 Mahé, F., Rognes, T., Quince, C., de Vargas, C., & Dunthorn, M. (2014). Swarm: robust
683 and fast clustering method for amplicon-based studies. *PeerJ*, *2*, e593.
684 doi:10.7717/peerj.593
685 Mandelik, Y., Roll, U., & Fleischer, A. (2010). Cost-efficiency of biodiversity indicators
686 for Mediterranean ecosystems and the effects of socio-economic factors.
687 *Journal of Applied Ecology*, *47*(6), 1179–1188. doi:10.1111/j.1365-
688 2664.2010.01864.x
689 Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput
690 sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. doi:10.14806/ej.17.1.200
691 Martínez, J., & Adarraga, I. (2006). Programa de vigilancia y control de la introducción
692 de especies invasoras en los ecosistemas litorales de la costa vasca. 1. *Costa de*
693 *Gipuzkoa*, *2*.
694 McMurdie, P. J. & Holmes, S. (2014). Waste not, Want Not: Why Rarefying Microbiome
695 Data is Inadmissible. P*los Computational Biology 10(4)*, e1003531.

Meyer, C. P. (2003). Molecular systematics of cowries (Gastropoda: Cypraeidae) and diversification patterns in the tropics. *Biological Journal of the Linnean Society*, *79*(3), 401–459. doi:10.1046/j.1095-8312.2003.00197.x

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

Rey, A., Basurko, O. C., & Rodríguez-Ezpeleta, N. (2018). The challenges and promises of genetic approaches for ballast water management. *Journal of Sea Research*, *133*, 134–145. doi:10.1016/j.seares.2017.06.001

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*

743 *and Environmental Microbiology*, *73*(16), 5261–5267. doi:10.1128/AEM.00062-
744 07
745 Xiong, W., Li, H., & Zhan, A. (2016). Early detection of invasive species in marine
746 ecosystems using high-throughput sequencing: technical challenges and
747 possible solutions. *Marine Biology*, *163*(6), 1–12.
748 Zaiko, A., Schimanski, K., Pochon, X., Hopkins, G. A., Goldstien, S., Floerl, O., & Wood, S.
749 A. (2016). Metabarcoding improves detection of eukaryotes from early
750 biofouling communities: implications for pest monitoring and pathway
751 management. *Biofouling*, *32*(6), 671–684.
752 doi:10.1080/08927014.2016.1186165
753 Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: a fast and accurate
754 Illumina Paired-End reAd mergeR. *Bioinformatics*, *30*(5), 614–620.
755 doi:10.1093/bioinformatics/btt593
756 Zorita, I., Solaun, O., Borja, A., Franco, J., Muxika, I., & Pascual, M. (2013). Spatial
757 distribution and temporal trends of soft-bottom marine benthic alien species
758 collected during the period 1989–2008 in the Nervión estuary (southeastern
759 Bay of Biscay). *Journal of Sea Research*, *83*, 104–110.
760
761

762 **Figures legends**

763

764 Figure 1. Sampling protocol representing sampling location (Port of Bilbao), sites (1, 2,
765 3 and 4) and points per site (dots); and illustrating the sampling methods used and the
766 targeted biological communities.

767

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

773

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

777

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

781

782 Figure 5. Principal Component Analyses of Hellinger-transformed abundances of OTUs
783 included in the *PBBS targeted taxa dataset* for COI (A-B) and 18S (C-D). A & C: Samples
784 scores in scaling 1 with ellipses representing the 95% confidence dispersion of each
785 sampling method. B & D: OTU scores in scaling 1 with the circle of equilibrium

786 contribution. Only OTUs whose projected length in these two principal component
787 axes exceeding the value of equilibrium contribution are represented.
788
789 Figure 6. Seasonal variation of alpha and beta diversity for each sampling method with
790 COI and 18S. A: Total OTU richness recovered at each season. B: OTU richness unique
791 to each season. C: Proportion of the total OTU richness detected with one season. D:
792 Decomposition of between-season beta diversity into replacement and nestedness
793 components. E: Proportion of the total OTU richness detected with two seasons.
794
795 Figure 7. Spatial variation of alpha and beta diversity for each sampling method with
796 COI and 18S. A: Total OTU richness recovered at each site. B: OTU richness unique to
797 each site. C: Proportion of the total OTU richness detected with one site. D:
798 Decomposition of between-site beta diversity into replacement and nestedness
799 components
800
801

802 **Appendix figure legends**

803 Appendix 1: Picture of the port depicting location of each site sampled. A: Examples of
804 the different types of structures at each site. B: Fouling communities settled on plates
805 after 3 months. Picture frames are colored according to the site they were taken at.

806 Appendix 2: OTU accumulation curves per sample for COI and 18S barcodes.

807

808 Appendix 3: Number of filtered water reads assigned to each phylum for three percent
809 identity ranges when blasting the unclassified reads (*i.e.* reads not classified below the
810 phylum level) against NCBI. The 5 most abundant phyla per percent identity range are
811 represented, the remaining being grouped under "other".

812

813 Appendix 4: Principal Component Analyses of the Hellinger-transformed abundances
814 of OTU included in the PBBS targeted metazoan taxa dataset for 18S. A: Samples
815 scores in scaling 1 with ellipses representing the 95% confidence dispersion of each
816 sampling method. B: OTU scores in scaling 1 with the circle of equilibrium contribution.
817 Only OTUs whose projected length in these two principal component axes exceeding
818 the value of equilibrium contribution are represented.

819

820 Appendix 5: Percentage of the total OTU richness detected by each sampling method
821 for the 10 most diverse metazoan orders of the PBBS targeted taxa dataset for each
822 barcode.

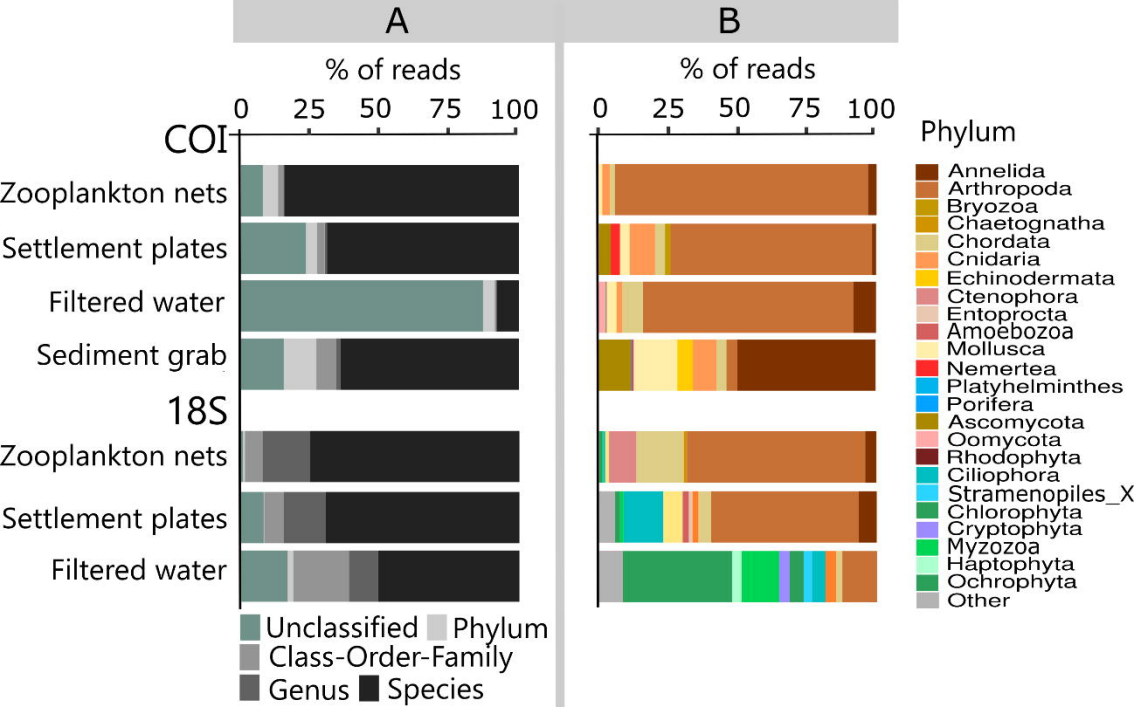
823

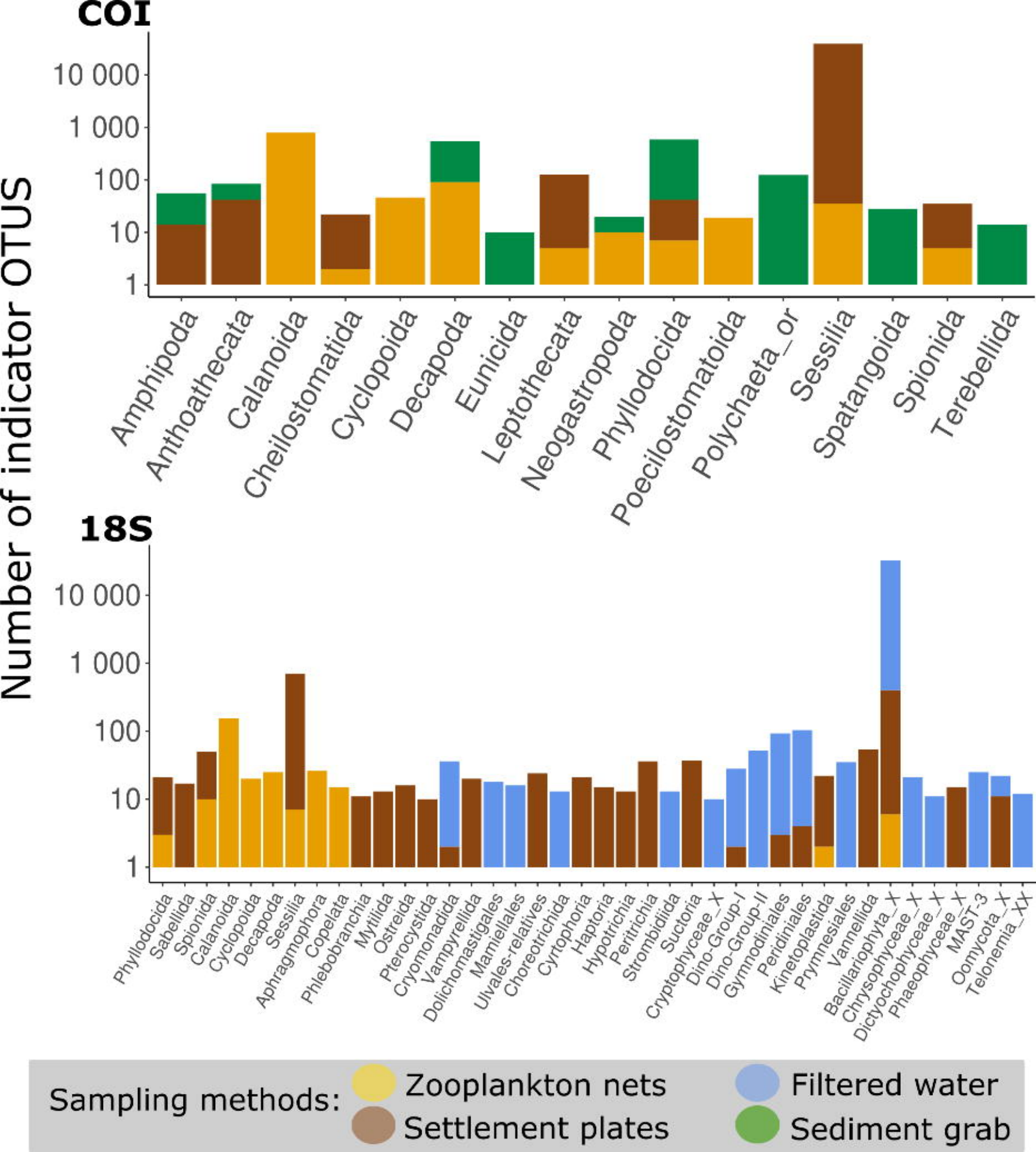
824 Appendix 6: Principal Component Analyses (PCA) of the Hellinger-transformed
825 abundances of OTUs included in the PBBS targeted taxa dataset of Vigo, A Coruña and
826 Bilbao filtered water samples. Sample scores are displayed in scaling 1. A: PCA
827 performed with 18S. B: PCA performed with 18S targeting only metazoan. C: PCA
828 performed with COI.

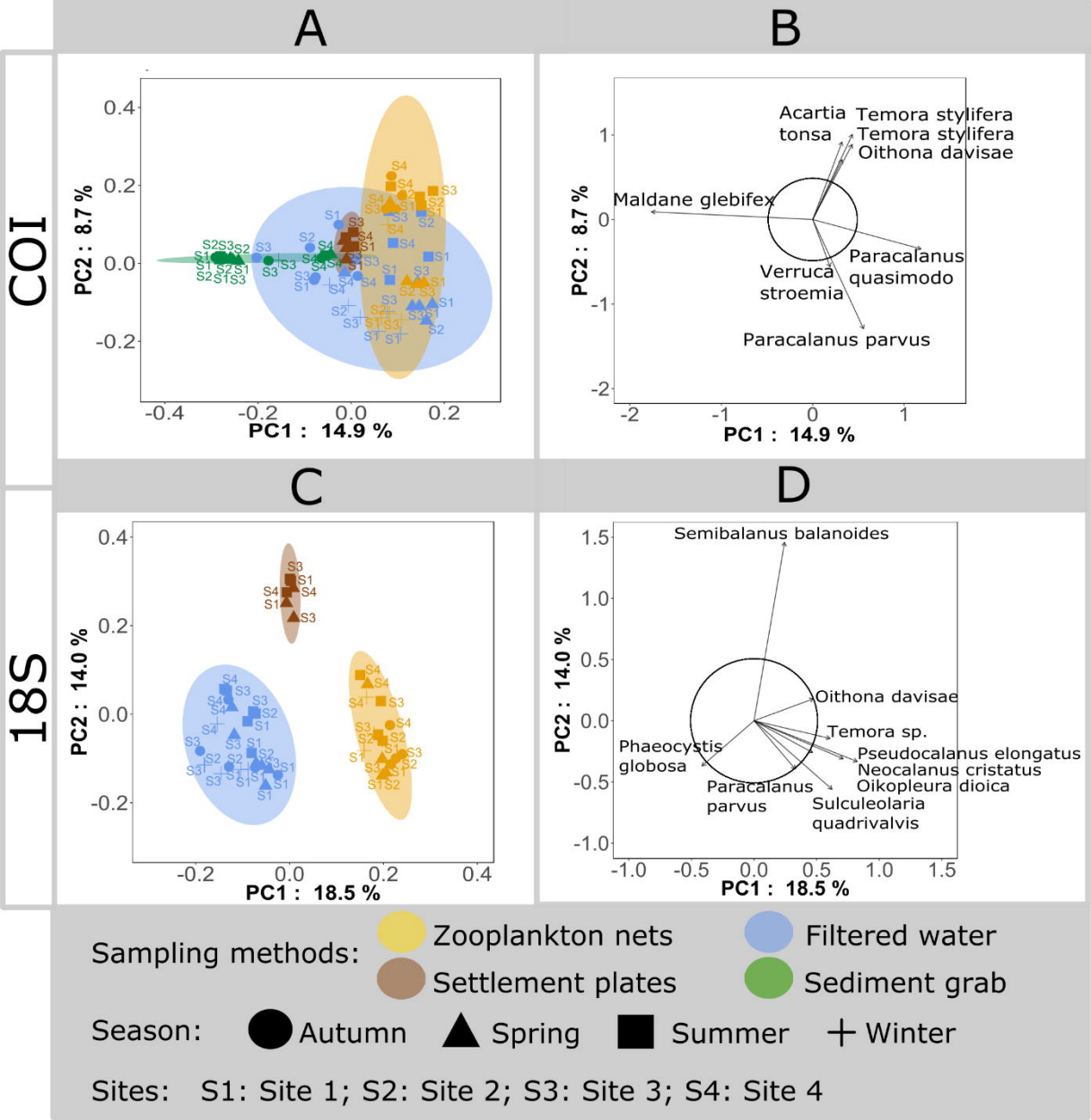
829

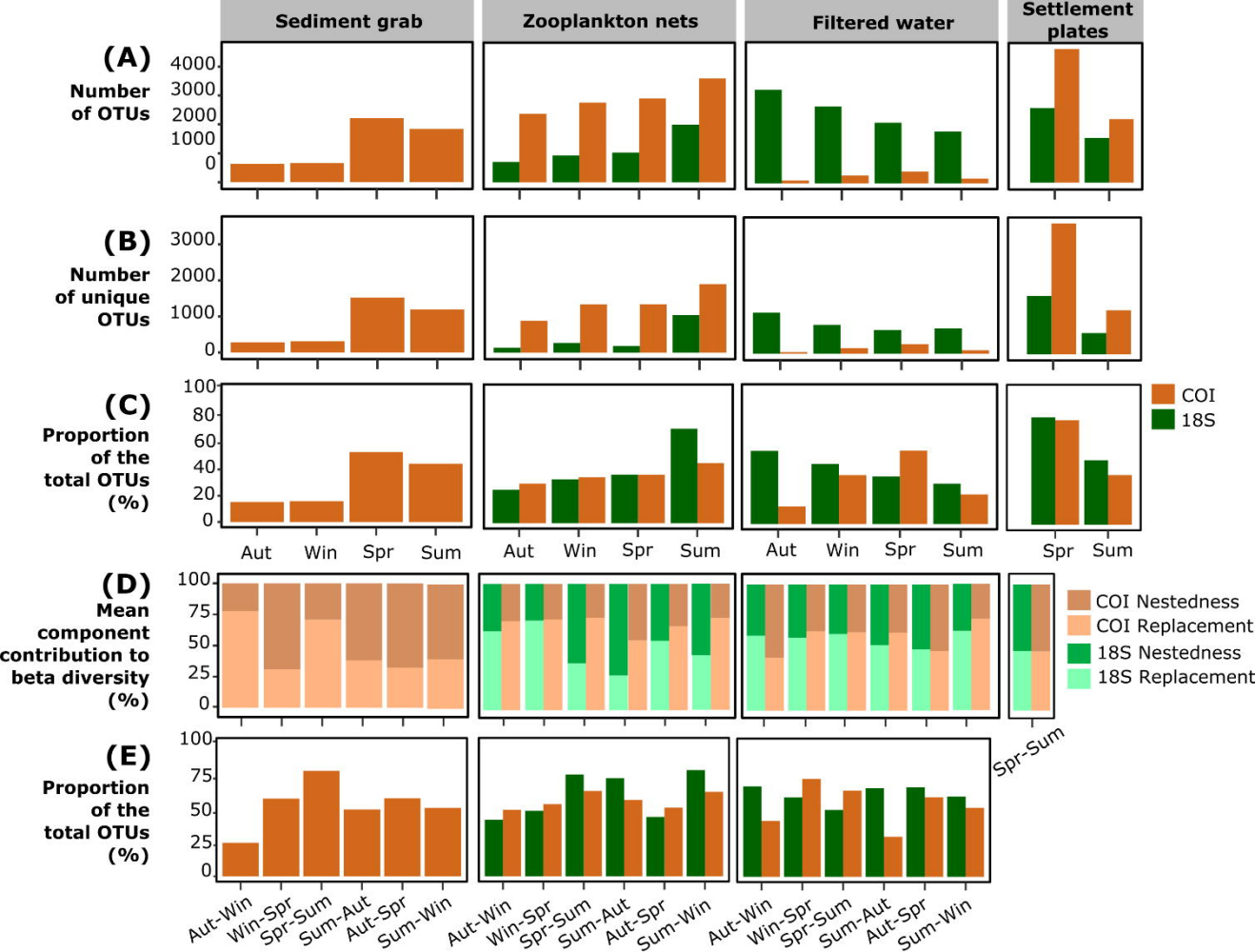
830

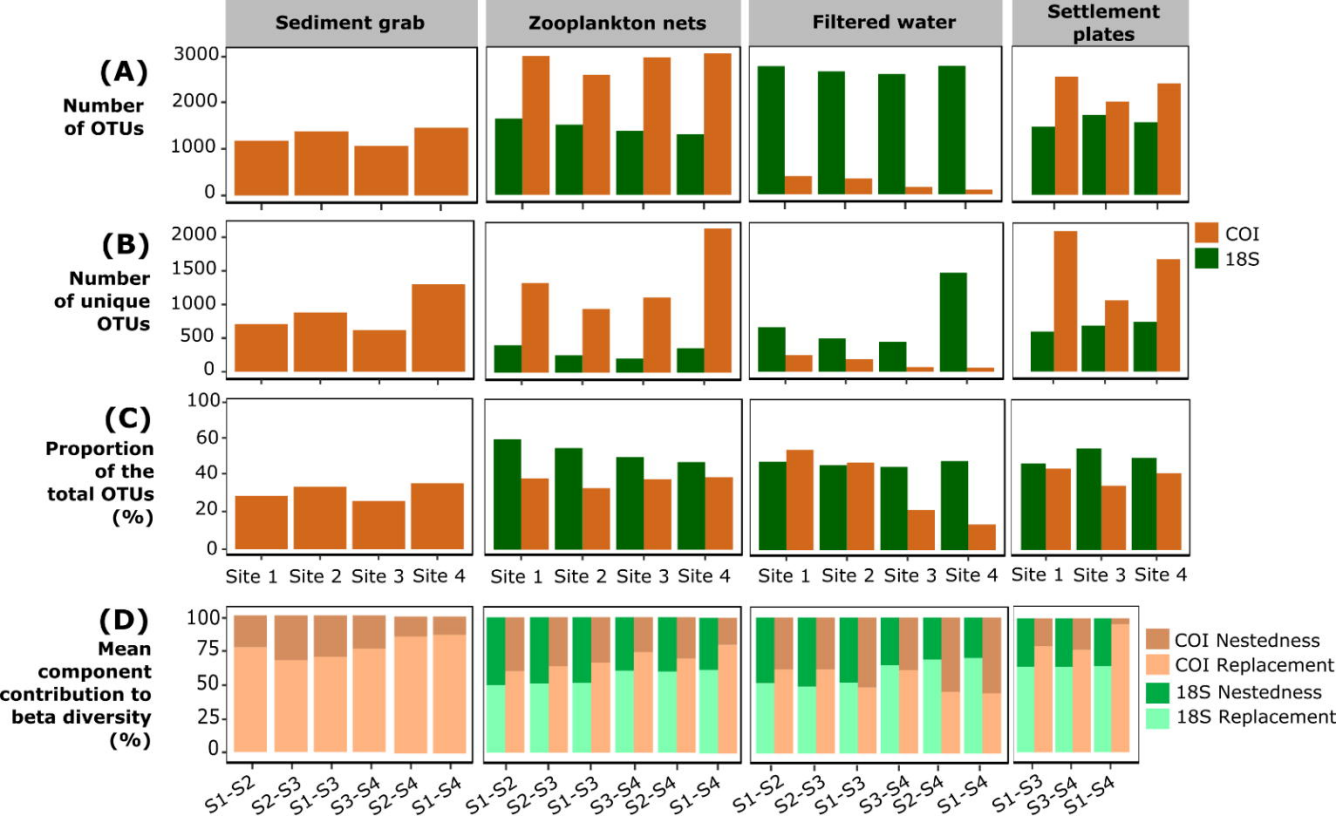
831

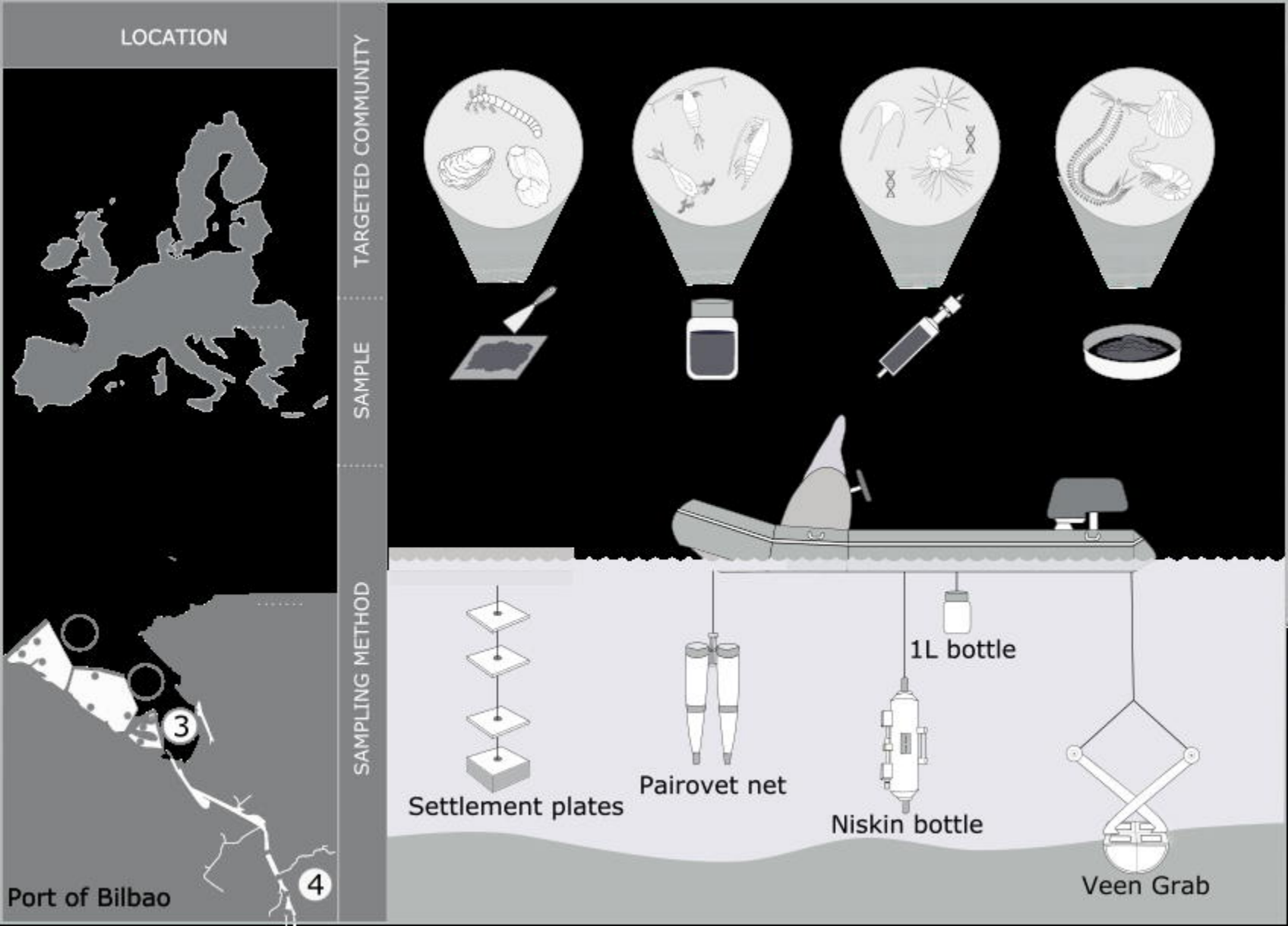

A
B
% of reads
0
25
50
75
100
COI
Zooplankton nets
Settlement plates
Filtered water
Sediment grab
18S
Zooplankton nets
Settlement plates
Filtered water
Unclassified
Phylum
Class-Order-Family
Genus
Species
Phylum
Annelida
Arthropoda
Bryozoa
Chaetognatha
Chordata
Cnidaria
Echinodermata
Ctenophora
Entoprocta
Amoebozoa
Mollusca
Nemertea
Platyhelminthes
Porifera
Ascomycota
Oomycota
Rhodophyta
Ciliophora
Stramenopiles_X
Chlorophyta
Cryptophyta
Myzozoa
Haptophyta
Ochrophyta
Other

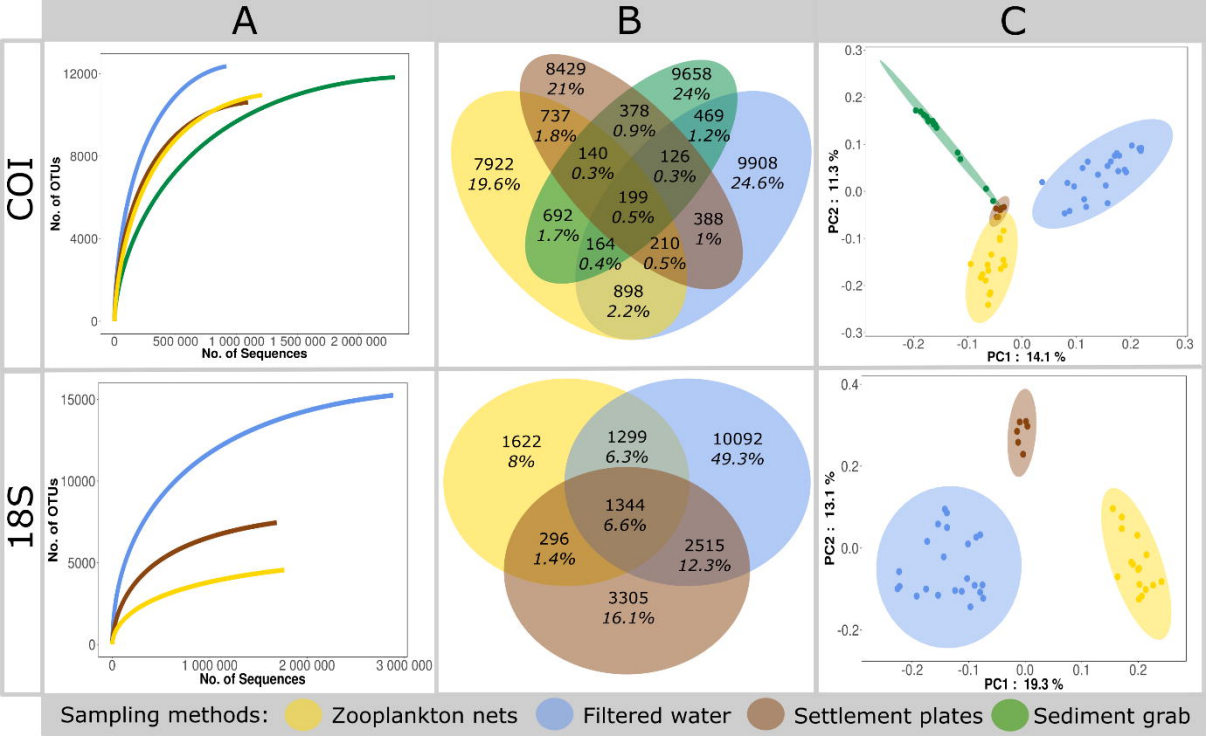

COI
18S
Number of indicator OTUS
10 000
1 000
100
10
1
Amphipoda
Anthoathecata
Calanoida
Cheilostomatida
Cyclopoida
Decapoda
Eunicida
Leptothecata
Neogastropoda
Phyllodocida
Poecilostomatoida
Polychaeta_or
Sessilia
Spatangoida
Spionida
Terebellida
Phyllodocida
Sabellida
Spionida
Calanoida
Cyclopoida
Decapoda
Sessilia
Aphragmophora
Copelata
Phlebobranchia
Mytilida
Ostreida
Pterocystida
Cryomonadida
Vampyrellida
Dolichomastigales
Mamiellales
Ulvales-relatives
Choreotrichida
Cyrtophoria
Haptoria
Hypotrichia
Peritrichia
Strombidiida
Suctoria
Cryptophyceae_X
Dino-Group-I
Dino-Group-II
Gymnodiniales
Peridiniales
Kinetoplastida
Prymnesiales
Vannellida
Bacillariophyta_X
Chrysophyceae_X
Dictyochophyceae_X
Phaeophyceae_X
MAST-3
Oomycota_X
Telonemia_XX
Sampling methods:
Zooplankton nets
Settlement plates
Filtered water
Sediment grab


A
B
C
D
COI
18S
PC1 : 14.9 %
PC2 : 8.7 %
Acartia tonsa
Temora stylifera
Temora stylifera
Oithona davisae
Maldane glebifex
Paracalanus quasimodo
Verruca stroemia
Paracalanus parvus
PC1 : 18.5 %
PC2 : 14.0 %
Semibalanus balanoides
Oithona davisae
Temora sp.
Phaeocystis globosa
Pseudocalanus elongatus
Neocalanus cristatus
Oikopleura dioica
Paracalanus parvus
Sulculeolaria quadrivalvis
Sampling methods:
Zooplankton nets
Filtered water
Settlement plates
Sediment grab
Season:
Autumn
Spring
Summer
Winter
Sites: S1: Site 1; S2: Site 2; S3: Site 3; S4: Site 4


Sediment grab
Zooplankton nets
Filtered water
Settlement plates
(A)
Number of OTUs
4000
3000
2000
1000
0
(B)
Number of unique OTUs
3000
2000
1000
0
(C)
Proportion of the total OTUs (%)
100
80
60
40
20
0
Aut
Win
Spr
Sum
COI
18S
(D)
Mean component contribution to beta diversity (%)
100
75
50
25
0
COI Nestedness
COI Replacement
18S Nestedness
18S Replacement
(E)
Proportion of the total OTUs (%)
100
75
50
25
0
Aut-Win
Win-Spr
Spr-Sum
Sum-Aut
Aut-Spr
Sum-Win


Sediment grab
Zooplankton nets
Filtered water
Settlement plates
(A)
Number of OTUs
(B)
Number of unique OTUs
(C)
Proportion of the total OTUs (%)
(D)
Mean component contribution to beta diversity (%)
COI
18S
Site 1 Site 2 Site 3 Site 4
COI Nestedness
COI Replacement
18S Nestedness
18S Replacement
S1-S2 S2-S3 S1-S3 S3-S4 S2-S4 S1-S4
S1-S3 S3-S4 S1-S4


LOCATION
TARGETED COMMUNITY
SAMPLE
SAMPLING METHOD
3
4
Port of Bilbao
Settlement plates
Pairovet net
1L bottle
Niskin bottle
Veen Grab


A
B
C
COI
18S
No. of OTUs
No. of Sequences
PC1 : 14.1 %
PC2 : 11.3 %
PC1 : 19.3 %
PC2 : 13.1 %
8429
21%
9658
24%
737
1.8%
378
0.9%
469
1.2%
7922
19.6%
140
0.3%
126
0.3%
9908
24.6%
199
0.5%
692
1.7%
388
1%
164
0.4%
210
0.5%
898
2.2%
1622
8%
1299
6.3%
10092
49.3%
1344
6.6%
296
1.4%
2515
12.3%
3305
16.1%
Sampling methods:
Zooplankton nets
Filtered water
Settlement plates
Sediment grab