Title

Metabarcoding hyperdiverse marine communities in temperate kelp forests: an experimental approach to inform future studies.

Short running title

cDNA metabarcoding in temperate kelp forests

Authors

Vanessa Arranz †[1], Libby Liggins[1] and J. David Aguirre[1]

[1]School of Natural and Computational Sciences, Massey University Auckland, Albany, Auckland, 0745, New Zealand.

†Corresponding author: vanearranz@hotmail.com

Abstract

Classical taxonomic approaches to quantifying biodiversity can be notoriously laborious and restrictive. Instead, molecular metabarcoding is emerging as a rapid, high-throughput and cost-effective tool to catalogue biodiversity. Despite the appeal of metabarcoding however, methodological and procedural biases must be understood before robust biodiversity inferences can be made. Here, we use CO1 metabarcoding to characterize marine eukaryote communities associated with the holdfasts of *Ecklonia radiata*, the dominant eco-engineering kelp of temperate New Zealand and Australia. To establish a standardized and reproducible community metabarcoding protocol, we examined the influence of different sample preparation, laboratory and bioinformatic steps on inferences of species richness and composition for kelp-holdfast communities. Specifically, we examined: the effect of fractioning the community into different size classes, the replicability of results across DNA extractions, PCR reactions, and sequencing. Overall, our approach identified 18 marine eukaryote Phyla in the holdfast communities. We found that size fractioning the sample before DNA extraction enabled detection of a greater diversity of taxa, especially smaller organisms. When compared with traditional morphology-based inventories of kelp-holdfast biodiversity, we found that although the taxonomic precision of our metabarcoding approach at the species and genus level was limited by the availability of reference sequences in public repositories, we recovered a greater number of operational taxonomic units, and a greater taxonomic breadth of organisms than morphological surveys. Based on our findings, we provide methodological guidelines for the use of metabarcoding as a tool for surveying and monitoring the hyperdiverse species assemblages associated with kelp-holdfasts.

Keywords

DNA metabarcoding, cDNA, Marine biodiversity, Monitoring, Kelp-holdfast assemblages, Kelp forests

Introduction

DNA metabarcoding (Taberlet, Coissac, Hajibabaei, & Rieseberg, 2012) has revolutionized the way we characterize biodiversity (Bush et al., 2019; Stat et al., 2017) as well as the assessment of ecosystem and environmental health (Aylagas, Borja, Irigoien, & Rodríguez-Ezpeleta, 2016). DNA metabarcoding methods are now used in empirical ecology (Harms-Tuohy, Schizas, & Appeldoorn, 2016), invasion biology (Thomas et al., 2019), biomonitoring and for conservation management (Barnes & Turner, 2016). Where the study focus is a specific community, or when a bulk specimen mixture is taken from a focal environment, we refer to these approaches as community DNA metabarcoding (cDNA; Creer et al., 2016). cDNA studies use the same methods of high-throughput DNA extraction, PCR and sequencing common to all metabarcoding approaches, but aim to directly identify the taxa within the sampled community based on their DNA barcode. These cDNA studies are akin to traditional visual, morphology-based surveys in aiming to characterize the taxonomic richness of a community or species assemblage, and to infer differences in the taxonomic composition among sampled communities (Deiner et al., 2017; Taberlet, Coissac, Pompanon, Brochmann, & Willerslev, 2012).

DNA-based monitoring methodologies (Baird & Hajibabaei, 2012), such as cDNA, have provided comparable results to traditional biodiversity surveys in a range of ecosystems (Deiner et al., 2017). In contrast to morphology-based surveys however, cDNA does not depend on expert taxonomic training (Bush et al., 2019), allows higher comparability across studies (Aylagas et al., 2016; Ji et al., 2013), and it produces data

64 (i.e. sequence reads) that can be easily re-analyzed and re-interpreted by a secondary
65 user. As a result, cDNA approaches often discover greater numbers of taxa within a
66 community than has been previously described (Bush et al., 2019; Siegenthaler,
67 Wangensteen, Benvenuto, Campos, & Mariani, 2019; Valentini et al., 2016). In the last
68 decade, cDNA approaches have been successfully used to describe past and present
69 biodiversity in terrestrial (Brehm et al., 2016; Dopheide et al., 2019), freshwater
70 (Andújar, Arribas, Yu, Vogler, & Emerson, 2018; Blackman et al., 2019; Elbrecht &
71 Leese, 2017; Hajibabaei, Porter, Wright, & Rudar, 2019), estuarine (Lobo, Shokralla,
72 Costa, Hajibabaei, & Costa, 2017) and marine (Aylagas et al., 2016; Knowlton & Leray,
73 2015; Zhang, Chain, Abbott, & Cristescu, 2018) environments. There is growing
74 recognition that cDNA can help characterize and monitor the biodiversity of all of earth's
75 important ecosystems and assist in making informed management decisions.

76 As an emerging tool however, before cDNA metabarcoding can be confidently
77 applied, rigorous examination of the potential biases and artefacts of the approach must
78 be conducted. For instance, several studies have addressed the influence of community
79 sampling protocols (Dickie et al., 2018), and laboratory methods, such as DNA
80 extraction procedures (Deiner et al., 2017; Lear et al., 2018), primer choice (van der
81 Loos & Nijland, 2020) and amplification bias (Kelly, Shelton, & Gallego, 2019), as well
82 as the level of replication at each methodological step, in producing robust, consistent
83 and reproducible results (Ficetola et al., 2015; Nichols et al., 2020; Porter et al., 2019).
84 Bioinformatic pipelines, which transform the sequence reads into community data, can
85 strongly influence study results and are constantly revised and improved (Pauvert et al.,

2019). Decisions within the bioinformatic pipelines, regarding the filtering of reads, processing PCR replicates, and sequencing depth have also been demonstrated to influence biodiversity estimates (Alberdi, Aizpurua, Gilbert, & Bohmann, 2018; Bokulich et al., 2013; Flynn, Brown, Chain, MacIsaac, & Cristescu, 2015; Kunin, Engelbrektson, Ochman, & Hugenholtz, 2010). As a consequence, prior to using a cDNA approach in a new ecosystem or focal community, there is a recognized need for experimental examination of the potential drawbacks and biases that different steps in the overall approach might introduce (Aylagas et al., 2016; Bush et al., 2019; McGee, Robinson, & Hajibabaei, 2019).

One of the most important considerations when first applying a cDNA approach in a new community, is establishing how to obtain a community DNA sample that is representative of the biodiversity present in a target community (Koziol et al., 2019). While it is common for procedural replicates and controls to be considered in the laboratory steps of cDNA, there have been few systematic examinations of the community sampling procedures on the overall estimates of biodiversity (Alberdi et al., 2018; Porter et al., 2019). Replicate samples of a target community, whether it be in traditional morphology-based biodiversity assessments or cDNA approaches, only recover a subset of the community, i.e. they are not a census, and as such the magnitude of among replicate variation must be quantified (Vlek, Šporka, & Krno, 2006). cDNA approaches generate millions of sequences and are therefore potentially able to reach the asymptote of the species discovery curve with fewer replicates than traditional surveys (Bush et al., 2019). However, this benefit of a cDNA approach will depend on

how representative each replicate DNA sample is of the community, and results vary among studies (Ficetola et al., 2015). It has been established that in communities where species have varying biomass, size fractioning of the community prior to DNA extraction can reduce misidentification or the omission of smaller organisms (Aylagas et al., 2016; Wangensteen, Cebrian, Palacín, & Turon, 2018). However, there has been little examination of subsequent steps in a cDNA approach, for instance, at what laboratory, or bioinformatic stage it is best to combine the different size fractions to recover representative estimates of biodiversity? Addressing the implications of these procedural decisions is important (Cowart et al., 2015); particularly, when the communities being characterized are known to support diverse taxa of varying sizes.

Kelp (Laminariales) are ecosystem-engineers (Jones, 2014) responsible for supporting incredibly diverse, structurally complex and highly productive ecosystems along temperate and polar coastlines worldwide (Steneck et al., 2002). On the Great Southern Reef of Australia, it has been estimated that between 700-4000 different species of algae, invertebrates and fishes occupy these ecosystems, with high levels of endemism (between 20-60%; Bennett et al. 2016). Accordingly, when these kelp are lost, we observe dramatic declines in biodiversity and ecosystem productivity (Bennett et al., 2016; Filbee-Dexter & Scheibling, 2014; Krumhansl et al., 2016; Ling, Johnson, Frusher, & Ridgway, 2009). Importantly, kelp are sensitive to environmental changes, and the demographic responses of kelp populations to stressors ripple throughout the ecosystem (Smale, Burrows, Moore, O'Connor, & Hawkins, 2013; Teagle, Hawkins, Moore, & Smale, 2017; Vergés et al., 2014; Wernberg et al., 2016). Monitoring

biodiversity changes within kelp forests using traditional survey methods is time-intensive and is highly dependent on scientists having diverse taxonomic expertise. For these reasons, the monitoring of kelp forest associated biodiversity has benefited from the use of metabarcoding approaches. Specifically, the analysis of environmental (e)DNA in seawater samples to detect vertebrate taxa within kelp forest ecosystems has gained results comparable to visual surveys (Port et al., 2016); and cDNA has also been used to characterize the sessile invertebrate communities attached to cobbles beneath the kelp canopy (Shum, Barney, O'Leary, & Palumbi, 2019).

The kelp holdfast is the structure which anchors the kelp to the substratum, and the complex web of haptera (root-like projections) which forms the holdfast provides a biogenically complex structure for a diversity of taxa and functional groups to colonize. Kelp holdfasts provide a logistically-convenient, biologically defined sampling unit which captures a broad diversity of marine eukaryote phyla (Anderson, Diebel, Blom, & Landers, 2005; Teagle et al., 2017). Despite the notable appeal of using kelp-holdfast assemblages as a barometer for change in this ecosystem, the large number, taxonomic diversity, and predominance of soft-bodied organisms, have precluded their use in morphological assessments of the biodiversity at the broad spatial scales necessary to monitor these ecosystems. Nonetheless, kelp-holdfasts could provide the basis for an effective cDNA approach to characterize and monitor biodiversity in kelp forest ecosystems.

Here, we evaluate metabarcoding as a tool for assessing biodiversity in kelp forest ecosystems using holdfast community DNA. Our comprehensive experimental

152 design examines the influence of methodological steps and decisions on biodiversity
153 estimates in these communities. Specifically, we assessed the effect of size fractioning
154 given the disparate sizes of organisms found within holdfast communities. We examined
155 the similarity between replicate samples of the same community to assess the
156 representativeness of each DNA extraction replicate and each PCR amplification of a
157 DNA extraction. Additionally, we analyzed the effect of bioinformatic decisions on final
158 biodiversity estimates. We then compare our taxonomically assigned cDNA reads with a
159 morphology-based, kelp-holdfast inventory from the same location to evaluate the
160 biases and opportunities of cDNA approach. Finally, we provide a series of guidelines
161 for community sampling, sample preparation, laboratory procedures and bioinformatic
162 decisions for metabarcoding marine eukaryotes from kelp-holdfast assemblages for
163 biodiversity surveys and monitoring.

164

165 **Materials and Methods**

166 **Experimental design**

167 Our overall goal was to assess the potential shortcomings and biases introduced by
168 different laboratory and bioinformatic procedures in the DNA metabarcoding of *Ecklonia*
169 *radiata* holdfast-associated communities. The impact of alternative decisions and
170 procedures on measures of taxonomic richness and community composition (hereafter
171 'biodiversity estimates') were examined at different stages of the cDNA workflow

172 including: sample preparation, DNA extraction, PCR amplification, and bioinformatic
173 manipulation (Fig. 1).

174 For sample preparation, we were interested in examining differences in biodiversity
175 estimates between the large (Lrg) and small (Sml) size fraction of the community
176 sample. Our experimental design aimed to test the influence of manually-pooling the
177 size fractions before DNA extraction (MPF_{COM}), manually-pooling the DNA extractions of
178 the large and small fraction before PCR (MPF_{EXT}), and manually-pooling the PCR
179 products of the large and small fraction before sequencing (MPF_{PCR}), as well as
180 bioinformatically-averaging (BAF) and bioinformatically-combining (BCF) the sequence
181 data of the large and small fraction on our biodiversity estimates (Fig. 1; for more detail
182 on the bioinformatic procedures, see Bioinformatic analysis section below). Specifically,
183 the contrasts we were interested in were:

184 1) Lrg vs. Sml: Do biodiversity estimates of the large and small size fractions differ?

185 2) BAF vs. MPF_{PCR} : Do biodiversity estimates differ if DNA extractions of the large
186 and small size fractions are PCRed and sequenced separately and then
187 bioinformatically-averaged, or if the large and small size fraction are manually-
188 pooled before sequencing?

189 3) BCF vs. MPF_{PCR} : Do biodiversity estimates differ if DNA extractions of the large
190 and small size fractions are PCRed and sequenced separately and then
191 bioinformatically-combined, or if the large and small size fractions are manually-
192 pooled before sequencing?

4) BAF vs. MPF_{EXT} : Do biodiversity estimates differ if DNA extractions of the large and small size fractions are PCRed and sequenced separately and then bioinformatically-averaged, or if the large and small size fraction are manually-pooled before PCR?

5) MPF_{EXT} vs. MPF_{COM} : Do biodiversity estimates differ if the PCR products of the large and small size fractions are manually-pooled before PCR, or if the large and small size fraction are manually-pooled before DNA extraction?

For DNA extraction, we examined differences between replicate extractions of the same community sample (Ext_A and Ext_B) and the influence of manually-pooling PCR products from the replicate extractions before sequencing (MPE_{PCR}), as well as bioinformatically-averaging (BAE) and bioinformatically-combining (BCE) sequence data for separately sequenced replicate extractions. Specifically, the contrasts we were interested in were:

6) Ext_A vs. Ext_B : Do biodiversity estimates of the two extraction replicates differ?

7) BAE vs. MPE_{PCR} : Do biodiversity estimates differ if the sequence data for the separately sequenced replicate extractions are bioinformatically-averaged, or if the PCR products of the replicate extractions are manually-pooled before sequencing?

8) BCE vs. MPE_{PCR} : Do biodiversity estimates differ if the sequence data for the separately sequenced replicate extractions are bioinformatically-combined, or if the PCR products of the replicate extractions are manually-pooled before sequencing?

For PCR amplification, we examined differences between replicate PCR products for the same extraction (PCR_X and PCR_Y) and the influence of manually-pooling replicate PCR

products for the same extraction before sequencing (MPP_{PCR}), as well as bioinformatically-averaging (BAP) and bioinformatically-combining (BCP) sequence data for replicate PCR products following sequencing. Specifically, the contrasts we were interested in were:

9) PCR_X vs. PCR_Y: Do biodiversity estimates from the two PCR product replicates differ?

10) BAP vs. MPP_{PCR}: Do biodiversity estimates differ if the sequence data for the separately sequenced replicate PCR products are bioinformatically-averaged, or if the PCR products of the replicate PCR product are manually-pooled before sequencing?

11) BCP vs. MPP_{PCR} : Do biodiversity estimates differ if the sequence data for the separately sequenced replicate PCR products are bioinformatically-combined, or if the PCR products of the replicate PCR product are manually-pooled before sequencing?

For the bioinformatic sample standardisation, we examined the influence of different filtering thresholds as well as the taxonomic precision of the sequence assignments (amplicon sequence variants [ASV] vs. Operational Taxonomic Units [OTU]) on our biodiversity estimates for the eleven contrasts described above (see Bioinformatic analysis section below for more detailed information).

Field sampling and processing

Mature *Ecklonia radiata* (mean height = 35.1cm range = 15.4 – 61.6cm) were collected from Mathesons Bay, Auckland, New Zealand by carefully sliding a knife between the base of the holdfast and the rocky reef. The individual kelps were quickly placed in polyethylene bags and sealed underwater to prevent organisms from escaping and the accidental transfer of organisms among samples. An additional seawater sample was collected so that the ambient environmental DNA could be separated from the holdfast cDNA. Upon returning to the surface, samples were placed in an insulated container and transported to the laboratory. In the laboratory, each *E. radiata* individual was placed in a separate tray and the holdfast community separated from the kelp-holdfast using forceps. The community sample was then passed through a stacked filter unit containing a fine, 63µm Sefar Nytal® filter on the bottom and a coarse 1000µm Sefar Nytal® filter on the top. The community sample retained on the 63µm filter was considered the small size fraction (Sml) containing meio-benthic organisms and the material retained on the 1000µm filter was considered the large size fraction (Lrg) containing mega- and macro-benthic organisms (Rex & Etter, 2010; Wangensteen et al., 2018). All sample processing equipment was sterilized prior to use on each holdfast by sterilizing the forceps with ethanol and flame, washing the filters and trays with soap and bleach, and by autoclaving the funnels. The two size fractions were transferred along with the mesh into individual, labelled 50mL falcon tubes and stored at -80ºC in preparation for lyophilization, more commonly known as freeze-drying. In the freeze-drying procedure, the falcon tubes containing the community samples were quickly transferred into a

Labconco FreeZone 6 bulk tray freeze-dryer with the lid off and they were subjected to a 24h cycle at a condensation temperature of -50ºC and a vacuum pressure of 0.12mbar. Once the freeze-drying was completed, a sterilized 15mL falcon tube was used as a pestle to grind the sample inside the 50mL falcon tube into a fine powder. The powdered community was then weighed and stored at -80ºC until DNA extraction.

Community DNA extraction and PCR

Genomic DNA was extracted from a 0.20g sample of the powdered kelp-holdfast community using the DNEasy PowerSoil Extraction Kit (QIAGEN) following the specified protocol, except using UltraPure DNase/RNAase-Free Distilled Water (Thermo Fisher) in the final step rather than elution buffer. For each sample, we performed two replicate extractions (Ext_A and Ext_B). For each set of extractions, a negative extraction control was also included using UltraPure water. To assess the quality and quantity of the extractions, the extracted DNA was run on a 1% agarose gel and visualized using GelRed® Nucleic acid gel stain (Biotium, Inc.), and gel-based estimates confirmed using a Qubit™ dsDNA br Assay Kit. The DNA concentration of all extractions was normalized to approximately 15ng/µL by diluting some samples with UltraPure DNase (the maximum required dilution was 1:10).

A 313 bp fragment of the Cytochrome Oxidase I (COI) mitochondrial gene was amplified using the primers mlCOIintF-XT (Wangensteen et al., 2018) and jgHCO2198 (Geller, Meyer, Parker, & Hawk, 2013), modified to include the Illumina™ overhang adaptors. Two PCR amplifications were performed for each DNA extract (PCR_X and

277 PCR_Y). For the PCR, 2.5µL of the DNA template was added to a mix consisting of
278 12.5µL of MyTaq™ Mix (Bioline, London), 1µL of each primer (10µM), 6µL of UltraPure
279 DNase/RNAase-Free Distilled Water (Thermo Fisher) and 2µL of Bovine Serum Albumin
280 (20mg/µL; Thermo Fisher) in a 25µL total reaction volume. We followed the Thermo-
281 cycling regime used by van der Reis, Laroche, Jeffs, and Lavery (2018) conducted in a
282 SureCycler 8800 Thermal Cycler (Agilent Technologies, USA). For each set of PCR
283 amplifications, a negative PCR control was also included using UltraPure water. PCR
284 amplicons were purified by magnetic separation following the Mag-Bind® Total Pure
285 NGS protocol (Omega, Bio-Tek). PCR products were pooled as required according to
286 our experimental design, quantified (Qubit® 2.0 Fluorometer, Invitrogen, Carlsbad, USA)
287 and diluted to equal concentration of between 5 and 15ng/µL. Sequencing was
288 performed at Massey Genome Services, Massey University (Palmerston North, New
289 Zealand) where indexing occurred using the Nextera™ DNA library Prep Kit (Illumina,
290 California, USA) before sequencing on an Illumina MiSeq™ System (2 × 250 paired-end
291 protocol).

292 **Bioinformatic analysis**

293 Sequence reads were analyzed and filtered using a series of quality control steps
294 available in the bioinformatics toolkit of QIIME 2 (Bolyen et al., 2019). First, the primers
295 were removed without mismatch tolerated using cutadapt (Martin, 2011). We used
296 DADA2 (Callahan et al., 2016) to perform the paired-end merging (trim-left r 13, trim-left-
297 f 13, trunc-len 200), dereplication, chimera filtering (using the consensus method) and
298 clustering of Amplicon Sequence Variants (ASV). The resulting ASV table was filtered

by a minimum abundance of reads across all samples using various thresholds to explore the effect of removing low abundance ASVs. The thresholds we examined were: no filtering, a minimum of 0.003% (Elbrecht & Leese, 2017), 0.01% (Alberdi et al., 2018) and 0.05% of the total read abundance across all samples. ASVs passing the quality control and filtering thresholds were taxonomically assigned using the MARES_COI_NOBAR reference sequence database (Arranz, Pearman, Aguirre, & Liggins, 2019). MARES is the most comprehensive COI reference database for marine eukaryotes available, and provides users the ability to retain taxa that cannot be assigned at the species level, but can be assigned at higher taxonomic levels – a desirable feature when working in communities of taxonomically diverse and potentially poorly characterized biodiversity (Arranz, Pearman, Aguirre, & Liggins, 2020). For taxonomic assignment, we first performed a BLASTn (Altschul, Gish, Miller, Myers, & Lipman, 1990) with an e-value of 1^{-60} for high-quality matches and max_target_seqs equal to 1. Then, we used MEGAN 6.18.3 (Huson et al., 2016), for taxonomic assignment within the NCBI taxonomy framework using the default Lowest Common Ancestor algorithm parameters. Using the decontam R package v1.4 (Davis, Proctor, Holmes, Relman, & Callahan, 2018) with the combined method, we filtered the ASV tables for contaminants found in the seawater eDNA sample as well as the extraction and PCR negative controls. Last, the ASVs were further clustered into Operational Taxonomic Units (OTUs) using VSEARCH v2.13.6 (Rognes, Flouri, Nichols, Quince, & Mahé, 2016) and a 97% similarity threshold.

Eight *E. radiata* holdfast communities were used in the experimental design, three of which had adequate biomass to be used in all the treatments used to construct the contrasts above (Fig. 1). This partially nested experimental design resulted in a total of 73 samples for sequencing including one seawater sample, three extraction negative controls and five PCR negative controls. We bioinformatically-combined samples by combining the appropriate sample size fractions (BCF), extractions (BCE) or PCRs (BCP) according to our experimental design, generating 17 synthetic samples for comparison. We bioinformatically-averaged samples by adjusting the coefficients of the contrast matrix (see Statistical analysis section below for details) for the appropriate fractions (BAF), extractions (BAE) and PCRs (BAP). A total of 81 samples were used for statistical analysis examined at both the ASV- and OTU-level and for all filtering thresholds as well as sample rarefaction. We considered the trade-off of increasing the rarefaction threshold to retain a greater proportion of the sampled diversity at the expense of removing greater numbers of samples with low ASV and OTU richness, or decreasing the rarefaction threshold and retaining more samples at the expense of removing a greater proportion of the sampled diversity. Given that our experimental design focuses on treatment (sample) contrasts, we deemed it was most important to select a rarefaction threshold for each ASV and OTU table that retained the greatest number of samples (Fig. S1, Supporting information). We assessed the effect of variability introduced by the rarefaction procedure by repeating each analysis on three different rarefied datasets each started from a different random seed as well as an analysis of the non-rarefied data (Table S2, Supporting information). Two samples ($MTB11\_Ext_A\_MPP_{PCR}$ and $MTB20\_Ext_B\_MPP_{PCR}$) had low sequencing depth for

unknown reasons, and these were excluded when the data were rarefied to even depth. Rarefaction was performed using the R package Phyloseq v1.28 (McMurdie & Holmes, 2013).

Statistical analysis

We used linear models implemented in R v3.6.1 (R Core Team, 2019) to assess how our laboratory and bioinformatic decisions influenced biodiversity estimates for the holdfast communities. First, we calculated the presence or absence of each ASV or OTU in each rarefaction and filtering category for each sample. From these presence-absence matrices we calculated two response variables: the observed taxonomic richness for each sample, and differences in community composition based on Jaccard's dissimilarity between all pairwise combinations of samples. Second, due to the unbalanced and partially nested nature of our experimental design, to specifically test the eleven planned comparisons described above, we set up a dummy variable, which assigned each sample to its corresponding combination of levels for the sample preparation, DNA extraction, PCR amplification and bioinformatic steps. For example, a sample from the large fraction where the replicate DNA extractions were manually-pooled before PCR amplification and then PCRed directly would have been coded as: $LrgMPE_{PCR}PCR_{X}$. This categorical dummy variable had a total of 25 levels for the 79 samples considered. We then constructed a contrast matrix for the dummy variable which contained our planned independent comparisons. Each column of the contrast matrix corresponded to a particular planned comparison (see Experimental design section above), allowing us to test contrasts directly, without unnecessary subsetting

and thereby multiple testing of the same data. Samples involved in contrasts were assigned weights of either 1 or -1 depending on the levels being contrasted, except for bioinformatically-averaged samples that were assigned a weight of ±0.5 (Crawley, 2012). Not all eleven contrasts were orthogonal. Non-orthogonal contrasts are analogous to collinear predictors and can produce similar statistical issues (Quinn & Keough, 2002); hence, to mitigate these issues we used the inverse of the transposed contrast matrix to calculate the fixed effects design matrix. We also used the inverse of the transposed contrast matrix for the calculation of the estimated marginal means and standard errors using the emmeans package v1.4.8 (Lenth, Singmann, Love, Buerkner, & Herve, 2018). Last, because individual kelp-holdfast communities contributed to multiple levels of the dummy variable (i.e. repeated measures), for our analysis of taxonomic richness, we specified the individual kelp identifier as a random effect in a linear mixed model fitted using the *lmer* function in the lme4 package v1.1 (Bates, Maechler, Bolker, & Walker, 2014). To test the statistical significance of the contrasts from the linear mixed models we used the *summary* function from the lmerTest package v3.1 (Kuznetsova, Brockhoff, & Christensen, 2017). For the analysis of community composition, we used the *adonis2* function in the vegan package v2.6 (Oksanen et al., 2007) and determined the significance of our contrasts using 999 permutations under a reduced model. While the *adonis2* function does not allow fitting mixed models, to account for repeated measures we constrained permutations to only occur among samples with the same individual kelp identifier.

Comparison with morphology-based surveys

To assess the performance of the cDNA approach, we compared the taxonomically assigned OTUs in the eight MPP_{PCR} samples against a morphology-based survey conducted for nine *E. radiata* holdfasts collected <750m away in 2002 (Anderson et al., 2005). We chose to focus on the MPP_{PCR} samples because according to our results, this would be the preferred treatment of future samples as it retrieved similar taxonomic richness and community composition as BCP samples while minimizing the sequencing costs (see Results section below). Prior to comparison, we assessed whether the assigned taxa identified by metabarcoding were of exclusively marine or brackish origin using the *wormsbyname* function in worms package v0.2 (Holstein, 2018). Assigned OTUs, which were not exclusively marine, were identified using a custom R script and removed. If an OTU or morphologically identified taxa could only be confidently assigned at a high taxonomic level (i.e. identifiable only to Class or Order) the OTU or morphologically identified taxa was labelled as undefined in lower taxonomic levels. We then compared the absolute and relative number of OTUs and morphologically identified taxa for these two studies from the same geographical area synonymizing taxonomies according to the World Register of Marine Species (WoRMS) (Horton et al., 2019).

Results

Across our experimental design, Illumina sequencing produced 4,310,106 paired-end reads. After quality filtering (primer removal, denoise, paired-end assembly, dereplication and chimera removal) a total of 947,469 reads were retained with a modal

sequence length of 313bp and a mean sequence length of 319bp. Two kelp-holdfast community samples with less than 5,000 reads were removed for downstream analysis due to their low number of reads (MTB11_Ext_A_MPP_{PCR} and MTB20_Ext_B_MPP_{PCR}). The final dataset after removing negative controls and possible contaminants consisted of 7,234 ASVs, with an average of 14,862 reads per sample (range: 5,156-24,624). ASVs were clustered into OTUs at 97% similarity, producing 2,671 OTUs. Filtering ASVs by minimum read abundance was performed at three levels (0.003 and 0.01 and 0.05%) and the filtered reads clustered into OTUs (Table S1, Supporting information). Rarefaction curves indicated that most of the samples approached an asymptote in ASV and OTU richness, indicating that sampling effort was sufficient to produce a representative estimate of the biodiversity in the sampled community (Fig. S1, Supporting information).

Sample preparation

For the eight sampled communities, the larger size fraction (Lrg) had overall lower taxonomic richness than the smaller size fraction (Sml) for all ASVs and OTUs, except when read abundance was strongly filtered (0.01-0.05% of total read abundance across samples; Fig. 2, Fig. 3 and Table S2, Supporting information). However, the large size fraction had a similar taxonomic richness to the small size fraction when comparing only the taxonomically assigned ASVs and OTUs (Fig. 2, Fig. 3 and Table S2, Supporting information). Conversely, though not unexpectedly, the community composition of the two size fractions differed strongly for every rarefaction, filtering, clustering and

taxonomic assignment procedure used (Fig. 2, Fig. 4 and Table S2, Supporting information).

The bioinformatically-combined sequence reads from the separately sequenced size fractions (BCF) retrieved more taxonomic richness than manually-pooling samples before sequencing (MPF_{PCR}) based on ASVs (Fig. 2, Fig. 3 and Table S2, Supporting information). Interestingly, the taxonomic richness of the bioinformatically-combined fractions (BCF) and the manually-pooled fractions before sequencing (MPF_{PCR}) was similar when we clustered the ASVs into OTUs at 97% similarity, regardless of the level of filtering or if reads were taxonomically assigned (Fig. 2 and Table S2, Supporting information). The lowest taxonomic richness was found when the sequence reads of the two size fractions were bioinformatically-averaged (BAF; Fig. 2, Fig. 3 and Table S2, Supporting information). Nevertheless, although we detected differences in taxonomic richness between bioinformatically-combining and bioinformatically-averaging the two size fractions (manually or bioinformatically) we detected no significant differences in community composition among these two bioinformatic approaches (Fig. 2 and Table S2, Supporting information).

Manually-pooling the large and small size fractions before DNA extraction (MPF_{COM}) or before PCR (MPF_{EXT}) had little effect on estimates of taxonomic richness except when more stringent filtering was applied at the OTU-level, where manually-pooling the size fractions before PCR (MPF_{EXT}) recovered greater taxonomic richness than manually-pooling the large and small size fractions before DNA extraction (MPF_{COM}; Fig. 2 and Table S2, Supporting information). Conversely, we found significant

differences in community composition between the manually-pooled size fractions before DNA extraction (MPF_{COM}) and the manually-pooled size fractions before PCR (MPF_{EXT}; Fig. 2 and Table S2, Supporting information). The taxonomic richness of the manually-pooled size fractions before PCR (MPF_{EXT}) was similar to the taxonomic richness retrieved by bioinformatically-averaging the sequence data of the two size fractions (BAF; Fig. 3 and Table S2, Supporting information). Interestingly, all methods for pooling the size fractions after separately extracting the DNA of each fraction (i.e. manually-pooling extractions of the size fractions before PCR, MPF_{EXT}; pooling PCR products of the size fractions before sequencing, MPF_{PCR}; bioinformatically-averaging, BAF; and bioinformatically-combining, BCF; sequence data from each fraction) showed similar community composition (Fig. 2 and Table S2, Supporting information).

DNA extraction

The taxonomic richness of replicate extractions from the same sample (EXT_A and EXT_B) did not differ significantly for most of the clustering, rarefaction and filtering options we examined, except for ASVs when no filtering by minimum read abundance was applied (Fig. 2 and Table S2, Supporting information). Additionally, we found no significant differences in community composition among replicate extractions for most of the combinations of clustering, taxonomic assignment and the level of filtering (Fig. 2 and Table S2, Supporting information).

Across the sampled communities, the mean taxonomic richness of the bioinformatically-averaged replicate extractions (BAE) was lower than the manually-pooled extractions (MPE_{PCR}; Fig. 3 and Table S2, Supporting information). Moreover,

when weak filtering was applied to ASVs, the bioinformatically-combined sequence reads of the extraction replicates (BCE) showed higher taxonomic richness (Table S2, Supporting information). This result suggests that ASVs and OTUs may differ among sub-samples of the same community (i.e. EXT_A and EXT_B). Based on this contrast alone, we cannot determine whether biases have been introduced by PCR, and/or sequencing, rather than the sub-sampling of the community. However, the community composition of the bioinformatically-averaged (BAE) and combined (BCE) replicate extractions was similar to the community composition retrieved by manually-pooling extraction replicates (MPE_{PCR}; Fig. 2 and Table S2, Supporting information).

PCR amplification

Samples from the same extraction, PCRed and sequenced separately (PCR_X and PCR_Y) differed significantly in taxonomic richness in most of our contrasts, except when applying the most stringent filtering (Fig. 2 and Table S2, Supporting information). However, the community composition of the PCR replicates did not differ significantly in any of the contrasts we examined (Fig. 2 and Table S2, Supporting information).

We found no significant differences in taxonomic richness among the different strategies for pooling PCR replicates for all and assigned OTUs as well as assigned ASVs for most levels of filtering (Table S2, Supporting information). However, at the ASV-level using all reads, the highest taxonomic richness was found when bioinformatically-combining the sequence data of both size fractions (BCF), followed by manually-pooling PCR replicates (MPP_{PCR}), and the lowest taxonomic richness was found for bioinformatically-averaging PCR replicates (BAP; Fig. 3). The community

composition of the PCR replicates did not differ significantly at either the ASV- or OTU-level when only the taxonomically assigned or all reads were examined (Fig. 2 and Table S2, Supporting information).

Comparison with morphology-based surveys

The morphology-based survey of holdfast-associated biodiversity recorded 181 taxa belonging to 14 phyla, of which 121 taxa were identified to the species-level (Anderson et al., 2005; Fig 5). Our metabarcoding-based approach identified a total of 314 OTUs, representing 18 phyla; however, only 48 OTUs were assigned to the species-level with 43 unique species identified (Fig. 5 and Table S3, Supporting information). The number of assigned OTUs (i.e. taxa) was higher at lower taxonomic ranks for morphology-based surveys and the number of undefined taxa lower (Table S3, Supporting information). However, at higher taxonomic ranks (Class, Order and Phylum) the number of assigned OTUs with the metabarcoding approach exceeded that of the morphology-based surveys (Table S3, Supporting information).

The taxonomic overlap between the morphology- and metabarcoding-based survey was minimal at low taxonomic ranks, only two species and six genera were found in both survey methodologies, barely 1-3% of the total OTUs assigned to those levels (Fig. 5 and Table S3, Supporting information). The taxonomic overlap increased at higher taxonomic ranks, reaching 34% and 33% of taxa identified at the Class and Phylum level, respectively (Fig. 5). There were three Phyla found only in the visual surveys (Brachiopoda, Rhizopoda and Sipuncula), though these represented a small percentage (a combined 2.2%) of the total taxa found by the morphology-based

518 approach (Fig. 6). Seven Phyla were exclusively identified using metabarcoding,
519 including microeukaryotes and fungi (Myzozoa, Ascomycota and Oomycota),
520 Archaeplastida and Stramenopiles (Rhodophyta, Chlorophyta, Bacillariophyta and
521 Ochrophyta) though some of these Phyla were excluded from the morphology-based
522 surveys *a priori* (Fig. 6). Importantly, these microeukaryotic and fungal phyla
523 represented 43% of the total taxa found using the metabarcoding approach (Fig. 6).
524 Arthropods, Annelids, Porifera and Echinoderms were common in the morphology- and
525 metabarcoding-based survey (Fig. 6). Interestingly, Molluscs and Bryozoans which were
526 common in the morphology-based survey were scarce in metabarcoding-based survey
527 (Fig. 6).

528 **Discussion**

529 To routinely apply DNA metabarcoding of kelp-holdfast-associated biodiversity as a kelp
530 forest ecosystem monitoring tool, sampling and laboratory protocols must be optimized,
531 validated and standardized (Cowart et al., 2015; Elbrecht & Leese, 2017; Pawlowski et
532 al., 2018). Here, we present the analysis of a robust experimental design that quantifies
533 the impacts of various practical, laboratory and bioinformatic decisions made during a
534 cDNA approach to estimating biodiversity. Our overall aim was to highlight the
535 opportunity for using cDNA to assess the taxonomic richness and community
536 composition of assemblages living on and in the holdfasts of a dominant ecosystem
537 engineering kelp, and to identify any shortcomings and biases in such an approach. Our
538 results highlight that fractioning the community into similarly sized organisms enables
539 detection of a wider range of taxa, and replication of community DNA extraction, as well

as replicate PCR reactions help to capture the maximum taxonomic richness within a sample. When compared with traditional morphology-based approaches to quantifying biodiversity in kelp-holdfast communities, a cDNA approach recovers higher levels of taxonomic richness and a greater breadth of phyla. Nonetheless, as described in several other systems (Gauthier et al., 2019), incomplete reference sequence databases, remain a key factor limiting the potential of cDNA approaches to biodiversity assessment in this ecosystem. Below we outline our learnings, and discuss their implications for quantifying biodiversity, providing methodological and procedural recommendations for cDNA studies of kelp-holdfast-associated biodiversity.

Partitioning the kelp-holdfast community into two size fractions, corresponding to the mega- as well as macro-benthic organisms (Large [Lrg] fraction) and meio-benthic organisms (Small [Sml] fraction), allowed the detection of a wider diversity than what can be achieved without size fractioning. Within kelp-holdfasts, resident organisms vary considerably in their biomass (Anderson et al., 2005), from nematodes only micrometers in length to sponges or colonial ascidians that can dominate much of the available space within a holdfast. The composition of the community obtained when pooling the large and small size fractions after extraction, regardless of whether samples were pooled manually or bioinformatically (MPF_{EXT}, MPF_{PCR}, BAF and BCF), was different to that of the community characterized when both fractions were extracted together, simulating the un-fractioned sample of the same communities (MPF_{COM}) (Fig. 2). Higher numbers of DNA copies from larger organisms with greater biomass can hinder the detection of smaller organisms, and thus in the absence of size fractioning, higher sequencing effort

562 is required to detect small, low biomass organisms (Cowart et al., 2015). In our study,
563 the smaller size fraction had greater taxonomic richness than the larger size fraction at
564 the ASV- and OTU-level. However, because many of these ASVs and OTUs had low
565 read abundance, after applying more stringent filtering, differences in richness between
566 the large and small fraction disappeared. Similar to previous studies, our results suggest
567 that without size fractioning, it may be difficult to recover the presence of small
568 organisms in taxonomically diverse communities (Rex & Etter, 2010; Wangensteen et
569 al., 2018).

570 To avoid missing taxa, previous studies have suggested that multiple extractions
571 and amplifications of the same sample may be required (Ficetola et al., 2015). Across
572 our study design, the mass of the community subsample was kept consistent, optimizing
573 the ratio of sample mass to reagent volume for DNA extraction (as determined in pilot
574 studies) and enabling the use of each community sample across several experimental
575 treatments. Despite our efforts to homogenize the samples before subsampling, we
576 found that although extraction replicates (EXT_A vs. EXT_B) had similar richness they
577 differed in community composition. This result is consistent with previous studies on
578 animal taxa that have also found high variability among extraction replicates (Hermans,
579 Buckley, & Lear, 2018). In the case of kelp-holdfast community samples, the size of any
580 subsamples used for extraction may often be too small relative to the bulk community
581 sample to recover the full taxonomic breadth of organisms that are present (Deiner et
582 al., 2017). Future studies may wish to trial increasing the overall mass of subsamples
583 used in DNA extraction, to potentially gain more representative samples of the entire

584 community. Nonetheless, our results also highlight the value of having extraction
585 replicates (Zhou et al., 2011).

586 In contrast, studies focused on single phyla have shown that replicate extractions
587 are less important than PCR replicates in minimizing variability among samples (Ficetola
588 et al., 2015; Porter et al., 2019). PCR replicates are recommended as a procedure to
589 reduce the PCR stochasticity and maximize the detection of taxa (Leray & Knowlton,
590 2015). The downside of increasing the number of PCR replicates however, is the
591 increased cost and the risk of false positives by accumulating artefactual sequences
592 (Alberdi et al., 2018; Ficetola et al., 2015). In our study, we found that although PCR
593 replicates are presumed to have similar biases due to primer choice and the laboratory
594 protocols we used, inherent stochasticity in each PCR replicate slightly influenced the
595 taxonomic richness observed, but not the community composition of replicate PCRs.

596 Despite our efforts to rarefy samples that were bioinformatically-combined
597 following sequencing so that they were comparable to samples that were manually-
598 pooled before sequencing, in most cases this was not sufficient to make up for the
599 impact of increased sequencing depth. For instance, bioinformatically-combining the
600 large and small fractions of the samples (BCF) tended to produce higher taxonomic
601 richness than if the large and small fractions were manually-pooled before sequencing
602 (MPF_{PCR}), and the lowest taxonomic richness was always observed when
603 bioinformatically-averaging the large and small fractions of the samples (BAF). Similar
604 results were observed for other pooling strategies used for the extraction replicates,
605 where we observed significantly higher richness when bioinformatically-combining the

606 extraction replicates (BCE) or manually-pooling PCRs of the replicate extractions
607 (MPE_{PCR}) compared with bioinformatically-averaging extraction replicates (BAE). In
608 contrast, bioinformatically-combining PCR replicates caused little or no increase in
609 taxonomic richness, except at the ASV-level where, again, bioinformatically-combined
610 PCR replicates (BCP) had the highest taxonomic richness. The high similarity in the
611 community composition of PCR replicates (discussed above) may explain why there was
612 no increase in taxonomic richness when PCR replicates were bioinformatically-
613 combined, supporting findings of previous studies that suggest ecological inferences are
614 influenced most by sequencing depth rather than PCR stochasticity (Smith & Peay,
615 2014).

616 Across our experimental design, the greatest taxonomic richness was recovered
617 through bioinformatically-combining fractions, extractions and PCR replicates (BCF,
618 BCE and BCP, respectively). Procedures equivalent to our bioinformatically-combined
619 treatment have been shown to commonly recover the highest number of species (Leray
620 & Knowlton, 2017). Nonetheless, the risk of false positives is also increased by such
621 additive strategies – through increased introduction and amplification of contaminants,
622 as well as sequencing errors. For these reasons, more restrictive strategies such as
623 removing singletons, even doubletons and tripletons (Kunin et al., 2010), and stringent
624 filtering can be used to remove artefactual sequences at the expense of removing low
625 abundance true positives (Elbrecht & Leese, 2017; Flynn et al., 2015; Leray & Knowlton,
626 2017). In our case, we used a range of thresholds for the minimum read abundance
627 filtering across samples to remove rare ASVs or OTUs that may be erroneous

sequences or artefacts. By choosing a relative read abundance across samples, ASVs which may appear in low abundance in some samples but may be present in greater abundance in other samples will be retained as they are possible true positives (Leray & Knowlton, 2017). Even using the most stringent filtering at the ASV-level, bioinformatically-combined fractions (BCF) recovered the highest diversity. Interestingly, when increasing the level of filtering, the advantage of maximizing the number of taxa detected by bioinformatically-combined extractions (BCE) and PCR replicates (BCP) disappears and estimates of richness become similar to those obtained by manually-pooling replicates before sequencing (MPE_{PCR} and MPP_{PCR}, respectively).

The chosen methodological approach for a cDNA study will differ depending on whether a study is focused on ASVs or OTUs and whether the overall richness or different measures of diversity or turnover among samples are of interest. Recently, the use of ASVs instead of OTUs has been promoted because it improves the reusability, reproducibility and comprehensiveness of sampled biodiversity (Callahan, McMurdie, & Holmes, 2017). In our study, differences in taxonomic richness between the bioinformatically-combined samples and other pooling strategies diminished when clustering ASVs into OTUs. At the ASV-level, combining the size fractions, extractions or PCRs bioinformatically after sequencing (BCF, BCE and BCP) revealed higher taxonomic richness than any other strategies for pooling samples, either by bioinformatically-averaging (BAF) or manually-pooling fractions before PCR (MPF_{EXT}) or before sequencing (MPF_{PCR}). However, at the OTU-level, samples manually-pooled before sequencing (MPF_{PCR}, MPE_{PCR} and MPP_{PCR}) showed similar richness as

650 bioinformatically-combined samples (BCF, BCE and BCP) at a reduced sequencing
651 cost. Because differences in taxonomic richness between pooled samples are no longer
652 significant when similar sequences are collapsed into OTUs, nucleotide differences
653 (<3%) among divergent lineages of the same species or cryptic species, appeared to
654 drive differences observed at the ASV-level.

655 Interestingly, the composition inferred for our kelp-holdfast communities remained
656 similar regardless of the strategy used for pooling fractions, extractions and PCRs and
657 were consistent across ASVs and OTUs. Therefore, it appears that multivariate
658 descriptions of community composition, may be more robust to methodological or
659 procedural biases than univariate biodiversity indices. If the main objective is to retrieve
660 the greatest number of taxa possible, bioinformatically-combining the fractions after
661 sequencing separately would be recommended (BCF), especially at the ASV-level,
662 using moderate filtering by read abundance (e.g. 0.01%) to remove false positives.
663 Nevertheless, if there are limits to resources or the primary interest of a study is focused
664 at the OTU-level, pooling the PCR replicates of the extractions for the large and small
665 fractions before sequencing (MPE_{PCR}) would recover a similar community composition.
666 Such an approach would minimize sequencing costs, in favor of increasing the field
667 sampling effort, and thereby potentially the overall richness captured by the study
668 (Porter et al., 2019).

669 We chose to use COI as the barcode region for characterizing the kelp-holdfast
670 community because of its substantial representation in reference sequence repositories
671 (Porter & Hajibabaei, 2018), its broad taxonomic coverage, and it has been shown to

successfully discriminate among species (Andújar et al., 2018). However, our data and previous results suggest that reference databases are biased towards highly abundant macro-organisms (Wangensteen et al., 2018), and lack reference sequences for small and cryptic species which can also be a challenge to identify morphologically. The observed difference in taxonomic richness between bioinformatically-combined samples of the large and small fractions (BCF) and manually-pooling the large and small fractions before sequencing (MPF_{PCR}) diminished as our reliance on taxonomic precision increased (i.e. considering only assigned ASVs and OTUs). Despite efforts to generate DNA barcodes for specific taxa and locations (Carew et al., 2017; Morinière et al., 2019), DNA metabarcoding is still somewhat limited by incomplete reference databases (Curry, Gibson, Shokralla, Hajibabaei, & Baird, 2018; McGee et al., 2019). While reference databases continue to improve, taxonomy-free approaches (Apothéloz-Perret-Gentil et al., 2017; Maechler, Walser, & Altermatt, 2020) enable some important biodiversity inferences, albeit without the tangible links to community function and resilience that require knowledge of species (or OTUs or ASVs) identity and ecology.

Using molecular tools for biomonitoring diverse assemblages is becoming more common, often detecting higher diversity than conventional morphology-based approaches (Deiner et al., 2017). Our metabarcoding approach retrieved almost two times the number of OTUs identified using conventional morphological surveys of kelp-holdfasts and had broader taxonomic coverage (Anderson et al., 2005; Shum et al., 2019; Wernberg et al., 2019). Nonetheless, there were certain strengths unique to each approach. For example, a higher diversity of taxa was found using cDNA for some

694 groups, such as Porifera and Cnidaria, with morphological features that are not easily
695 retained through common preservation techniques (e.g. freezing or fixation). Our cDNA
696 approach also recovered Rhodophyta and Orchophyta, Phyla that were not considered
697 by the morphological survey despite being important components of kelp forest
698 communities (Shum et al., 2019). On the other hand, using the metabarcoding
699 approach, Molluscs and Bryozoans were poorly represented relative to the morphology-
700 based survey, potentially due the strong taxonomic expertise in these groups (Anderson
701 et al., 2005) or due to a true loss in diversity in those groups as the two studies were
702 conducted 20 years apart. It could also be due to biases in the extraction of DNA from
703 organisms in mixed communities (Hermans et al., 2018) or biases introduced by primer
704 choice (van der Loos & Nijland, 2020). In this case, using alternative extraction
705 techniques, or using a combination of primers, to target certain taxa within the holdfast
706 more specifically may be helpful (Alberdi et al., 2018).

707 Overall, the metabarcoding approach captured a good representation of the
708 known kelp-holdfast diversity and proved more time- and cost-effective. For the groups
709 considered by both survey methods, there were similar trends in the number of taxa
710 recovered for the dominant phyla. For instance, Arthropoda, Annelida, Porifera and
711 Echinodermata – all abundant and important taxa in kelp forest ecosystems (Anderson
712 et al., 2005; Wernberg et al., 2019) – were common in both surveys showing a high
713 proportion of taxa. However, there was a higher proportion of undefined taxa using the
714 metabarcoding approach, especially at lower taxonomic ranks. The limited ability to
715 taxonomically assign the molecular OTUs, especially at lower taxonomic ranks, again,

716 reflects gaps in reference sequence databases (Wangensteen et al., 2018), particularly
717 for marine species (Arranz et al., 2020; Leray & Knowlton, 2016). Ideally, both
718 approaches would be used in tandem to further develop an understanding of their
719 respective strengths and weaknesses, and to provide a specimen reference collection
720 corresponding to the sequence reference database, to help increase the ability of cDNA
721 studies to assign taxonomic identities to sequences.

722 The effectiveness of a biomonitoring strategy depends on the ability to detect
723 diversity and change over time and space (Shum et al., 2019). Our study retrieved a
724 high number of taxa within a relatively low number of samples and demonstrated the
725 ability to distinguish among samples at small spatial scales (meters apart; Fig. 4). The
726 holdfast has been a key focus in ecological studies because it is convenient to sample,
727 hosts a diversity of taxa, and because kelps are susceptible to environmental change,
728 so too are their holdfasts (Smale et al., 2013; Teagle et al., 2017; Vergés et al., 2014;
729 Wernberg et al., 2016). One of the drawbacks in using the holdfasts for monitoring
730 however, was the immense diversity they support, making morphological
731 characterization of their associated biodiversity a highly intensive task. Our study
732 reveals that community metabarcoding provides a means for the high-resolution
733 characterization of biodiversity associated with holdfasts, thus making kelp-holdfast
734 assemblages an accessible barometer for monitoring biodiversity change in critically
735 important and at-risk kelp forests. By carefully dissecting procedural sources of bias and
736 determining cost-effective and reproducible methods it shows promise that community

737 DNA metabarcoding could provide a standardized method for sampling the biodiversity
738 of these hyperdiverse marine communities.

739

740 **Acknowledgements**

741 We would like to thank Marie Wong, Negah Nikanjam and Nuri Begum for their
742 assistance in the use of the freeze-dryer machine at Massey University; Dave Wright
743 and Emma Betty for help in the field collection, and Chandler Skinner-Horne, Maddie
744 MacLean Stones, Jenny Ann Sweatman, Julia Kim, Hochang Yoo, Ella Lis, Beau
745 Masters, Carrie Reyden, Georgia Hrstich and Katie Littlewood for help with sample
746 preparation. We are grateful to Ulla Von Ammon, Xavier Turón, Owen Wangensteen,
747 Ramon Gallego, William Pearman and Aimee Van der Reis for sharing their knowledge
748 and providing advice; and Marti J. Anderson for providing advice and together with
749 Wilma Blom for sharing the morphology-based kelp-holdfast biodiversity data. Special
750 thanks to Seacology.co.nz and Allgenetic.eu for the images in Fig. 1. Vanessa Arranz
751 was supported by a Marsden Fund Fast-Start grant (14-MAU-077), managed by Royal
752 Society Te Apārangi grant awarded to J. David Aguirre.

753

References

Alberdi, A., Aizpurua, O., Gilbert, M. T. P., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution, 9*(1), 134-147.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology, 215*(3), 403-410.

Anderson, M. J., Diebel, C. E., Blom, W. M., & Landers, T. J. (2005). Consistency and variation in kelp holdfast assemblages: spatial patterns of biodiversity for the major phyla at different taxonomic resolutions. *Journal of Experimental Marine Biology and Ecology, 320*(1), 35-56.

Andújar, C., Arribas, P., Yu, D. W., Vogler, A. P., & Emerson, B. C. (2018). Why the COI barcode should be the community DNA metabarcode for the metazoa. *Mol. Ecol., 27*(20), 3968-3975. doi:10.1111/mec.14844

Apothéloz-Perret-Gentil, L., Cordonier, A., Straub, F., Iseli, J., Esling, P., & Pawlowski, J. (2017). Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring. *Molecular ecology resources, 17*(6), 1231-1242.

Arranz, V., Pearman, W. S., Aguirre, J. D., & Liggins, L. (2019). *MARES Custom Metabarcoding Database*. Retrieved from: http://doi.org/10.17605/osf.io/8rdqk

Arranz, V., Pearman, W. S., Aguirre, J. D., & Liggins, L. (2020). MARES, a replicable pipeline and curated reference database for marine eukaryote metabarcoding. *Scientific Data, 7*(1), 1-8.

Aylagas, E., Borja, Á., Irigoien, X., & Rodríguez-Ezpeleta, N. (2016). Benchmarking DNA metabarcoding for biodiversity-based monitoring and assessment. *Frontiers in Marine Science, 3*, 96.

Baird, D. J., & Hajibabaei, M. (2012). Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology, 21*(8), 2039-2044.

Barnes, M. A., & Turner, C. R. (2016). The ecology of environmental DNA and implications for conservation genetics. *Conservation genetics, 17*(1), 1-17.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. In.

Bennett, S., Wernberg, T., Connell, S. D., Hobday, A. J., Johnson, C. R., & Poloczanska, E. S. (2016). The 'Great Southern Reef': social, ecological and economic value of Australia's neglected kelp forests. *Marine and Freshwater Research, 67*(1), 47-56.

Blackman, R. C., Mächler, E., Altermatt, F., Arnold, A., Beja, P., Boets, P., . . . Jones, J. I. (2019). Advancing the use of molecular methods for routine freshwater macroinvertebrate biomonitoring–the need for calibration experiments. *Metabarcoding and Metagenomics, 3*, 49-57.

Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., . . . Caporaso, J. G. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature methods, 10*(1), 57-59.

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., . . . Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology, 37*(8), 852-857. doi:10.1038/s41587-019-0209-9

Brehm, G., Hebert, P. D., Colwell, R. K., Adams, M.-O., Bodner, F., Friedemann, K., . . . Fiedler, K. (2016). Turning up the heat on a hotspot: DNA barcodes reveal 80% more species of geometrid moths along an Andean elevational gradient. *PLoS One, 11*(3), e0150327.

Bush, A., Compson, Z. G., Monk, W., Porter, T. M., Steeves, R., Emilson, E. J., . . . Baird, D. (2019). Studying ecosystems with DNA metabarcoding: lessons from biomonitoring of aquatic macroinvertebrates. *Frontiers in Ecology and Evolution, 7*, 434.

Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME journal, 11*(12), 2639.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nature methods, 13*(7), 581-583.

Carew, M. E., Nichols, S. J., Batovska, J., St Clair, R., Murphy, N. P., Blacket, M. J., & Shackleton, M. E. (2017). A DNA barcode database of Australia's freshwater macroinvertebrate fauna. *Marine and Freshwater Research, 68*(10), 1788-1802. doi:https://doi.org/10.1071/MF16304

Cowart, D. A., Pinheiro, M., Mouchel, O., Maguer, M., Grall, J., Miné, J., & Arnaud-Haond, S. (2015). Metabarcoding is powerful yet still blind: a comparative analysis of morphological and molecular surveys of seagrass communities. *PLoS One, 10*(2), e0117562. doi:10.1371/journal.pone.0117562

Crawley, M. J. (2012). *The R book*: John Wiley & Sons.

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.

Curry, C. J., Gibson, J. F., Shokralla, S., Hajibabaei, M., & Baird, D. J. (2018). Identifying North American freshwater invertebrates using DNA barcodes: are existing COI sequence libraries fit for purpose? *Freshwater Science, 37*(1), 178-189.

Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A., & Callahan, B. J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome, 6*(1), 226.

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., . . . De Vere, N. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology, 26*(21), 5872-5895.

Dickie, I. A., Boyer, S., Buckley, H. L., Duncan, R. P., Gardner, P. P., Hogg, I. D., . . . Morales, S. E. (2018). Towards robust and repeatable sampling methods in eDNA-based studies. *Molecular ecology resources, 18*(5), 940-952.

Dopheide, A., Tooman, L. K., Grosser, S., Agabiti, B., Rhode, B., Xie, D., . . . Drummond, A. J. (2019). Estimating the biodiversity of terrestrial invertebrates on a forested island using DNA barcodes and metabarcoding data. *Ecological Applications, 29*(4), e01877.

Elbrecht, V., & Leese, F. (2017). Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment. *Frontiers in Environmental Science, 5*(11). doi:10.3389/fenvs.2017.00011

Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., . . . Pompanon, F. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular ecology resources, 15*(3), 543-556.

Filbee-Dexter, K., & Scheibling, R. E. (2014). Sea urchin barrens as alternative stable states of collapsed kelp ecosystems. *Marine ecology progress series, 495*, 1-25.

Flynn, J. M., Brown, E. A., Chain, F. J., MacIsaac, H. J., & Cristescu, M. E. (2015). Toward accurate molecular identification of species in complex environmental samples: testing the performance of sequence filtering and clustering methods. *Ecology and evolution, 5*(11), 2252-2266.

Gauthier, M., Konecny-Dupré, L., Nguyen, A., Elbrecht, V., Datry, T., Douady, C., & Lefébure, T. (2019). Enhancing DNA metabarcoding performance and applicability with bait capture enrichment and DNA from conservative ethanol. *Molecular ecology resources*.

Geller, J., Meyer, C., Parker, M., & Hawk, H. (2013). Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Molecular ecology resources, 13*(5), 851-861.

Hajibabaei, M., Porter, T. M., Wright, M., & Rudar, J. (2019). COI metabarcoding primer choice affects richness and recovery of indicator taxa in freshwater systems. *PLoS One, 14*(9).

Harms-Tuohy, C. A., Schizas, N. V., & Appeldoorn, R. S. (2016). Use of DNA metabarcoding for stomach content analysis in the invasive lionfish Pterois volitans in Puerto Rico. *Marine ecology progress series, 558*, 181-191.

Hermans, S. M., Buckley, H. L., & Lear, G. (2018). Optimal extraction methods for the simultaneous analysis of DNA from diverse organisms and sample types. *Molecular ecology resources, 18*(3), 557-569. doi:doi:10.1111/1755-0998.12762

Holstein, J. (2018). worms: Retriving aphia information from World Register of Marine Species. *R package version 0.2, 2*.

Horton, T., Kroh, A., Ahyong, S., Bailly, N., Boyko, C. B., Brandão, S. N., . . . Zhao, Z. (2019). World Register of Marine Species (WoRMS) (Publication no. 10.14284/170). Retrieved 2019-12-17, from WoRMS Editorial Board http://www.marinespecies.org

Huson, D. H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., . . . Tappu, R. (2016). MEGAN Community edition - Interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol., 12*(6), e1004957. doi:10.1371/journal.pcbi.1004957

Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., . . . Edwards, F. A. (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology letters, 16*(10), 1245-1257.

Jones, P. J. (2014). *Governing marine protected areas: resilience through diversity*: Routledge.

Kelly, R. P., Shelton, A. O., & Gallego, R. (2019). Understanding PCR processes to draw meaningful conclusions from environmental DNA studies. *Scientific reports, 9*(1), 1-14.

Knowlton, N., & Leray, M. (2015). Exploring Coral Reefs Using the Tools of Molecular Genetics. In C. Birkeland (Ed.), *Coral Reefs in the Anthropocene* (pp. 117-132). Dordrecht: Springer Netherlands.

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.

Krumhansl, K. A., Okamoto, D. K., Rassweiler, A., Novak, M., Bolton, J. J., Cavanaugh, K. C., . . . Ling, S. D. (2016). Global patterns of kelp forest change over the past half-century. *Proceedings of the National Academy of Sciences, 113*(48), 13785-13790.

Kunin, V., Engelbrektson, A., Ochman, H., & Hugenholtz, P. (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental microbiology, 12*(1), 118-123.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software, 82*(13), 1-26.

Lear, G., Dickie, I., Banks, J., Boyer, S., Buckley, H. L., Buckley, T. R., . . . Hermans, S. (2018). Methods for the extraction, storage, amplification and sequencing of DNA from environmental samples. *New Zealand Journal of Ecology, 42*(1), 10-50A.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Emmeans: Estimated marginal means, aka least-squares means. *R package version, 1*(1), 3.

Leray, M., & Knowlton, N. (2015). DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences, 112*(7), 2076-2081.

Leray, M., & Knowlton, N. (2016). Censusing marine eukaryotic diversity in the twenty-first century. *Philos. Trans. R. Soc. Lond. B Biol. Sci., 371*(1702). doi:10.1098/rstb.2015.0331

Leray, M., & Knowlton, N. (2017). Random sampling causes the low reproducibility of rare eukaryotic OTUs in Illumina COI metabarcoding. *PeerJ, 5*, e3006.

Ling, S., Johnson, C., Frusher, S., & Ridgway, K. (2009). Overfishing reduces resilience of kelp beds to climate-driven catastrophic phase shift. *Proceedings of the National Academy of Sciences, 106*(52), 22341-22345.

Lobo, J., Shokralla, S., Costa, M. H., Hajibabaei, M., & Costa, F. O. (2017). DNA metabarcoding for high-throughput monitoring of estuarine macrobenthic communities. *Scientific reports, 7*(1), 1-13.

Maechler, E., Walser, J.-C., & Altermatt, F. (2020). Decision making and best practices for taxonomy-free eDNA metabarcoding in biomonitoring using Hill numbers. *bioRxiv*.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal, 17*(1), 10-12.

McGee, K. M., Robinson, C., & Hajibabaei, M. (2019). Gaps in DNA-based biomonitoring across the globe. *Frontiers in Ecology and Evolution, 7*, 337.

McMurdie, P. J., & Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one, 8*(4), e61217.

Morinière, J., Balke, M., Doczkal, D., Geiger, M. F., Hardulak, L. A., Haszprunar, G., . . . Rulik, B. (2019). A DNA barcode library for 5,200 German flies and midges (Insecta: Diptera) and its implications for metabarcoding-based biomonitoring. *Molecular ecology resources, 19*(4), 900-928.

Nichols, S. J., Kefford, B. J., Campbell, C. D., Bylemans, J., Chandler, E., Bray, J. P., . . . Furlan, E. M. (2020). Towards routine DNA metabarcoding of macroinvertebrates using bulk samples for freshwater bioassessment: Effects of debris and storage conditions on the recovery of target taxa. *Freshwater Biology, 65*(4), 607-620.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M. H. H., Oksanen, M. J., & Suggests, M. (2007). The vegan package. *Community ecology package, 10*, 631-637.

Pauvert, C., Buée, M., Laval, V., Edel-Hermann, V., Fauchery, L., Gautier, A., . . . Vacher, C. (2019). Bioinformatics matters: the accuracy of plant and soil fungal community data is highly dependent on the metabarcoding pipeline. *Fungal Ecology, 41*, 23-33.

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

Port, J. A., O'Donnell, J. L., Romero-Maraccini, O. C., Leary, P. R., Litvin, S. Y., Nickols, K. J., . . . Kelly, R. P. (2016). Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular Ecology, 25*(2), 527-541.

Porter, T. M., & Hajibabaei, M. (2018). Over 2.5 million COI sequences in GenBank and growing. *PloS one, 13*(9), e0200177. doi:10.1371/journal.pone.0200177

Porter, T. M., Morris, D. M., Basiliko, N., Hajibabaei, M., Doucet, D., Bowman, S., . . . Wainio-Keizer, K. (2019). Variations in terrestrial arthropod DnA metabarcoding methods recovers robust beta diversity but variable richness and site indicators. *Scientific reports, 9*(1), 1-11.

Quinn, G., & Keough, M. (2002). Experimental design and data analysis for biologists Cambridge Cambridge University Press. 537.

R Core Team. (2019). R: A language and environment for statistical computing. (Version 3.6.1): R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org

Rex, M. A., & Etter, R. J. (2010). *Deep-sea biodiversity: pattern and scale*: Harvard University Press.

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ, 4*, e2584. doi:10.7717/peerj.2584

Shum, P., Barney, B. T., O'Leary, J. K., & Palumbi, S. R. (2019). Cobble community DNA as a tool to monitor patterns of biodiversity within kelp forest ecosystems. *Molecular ecology resources, 19*(6), 1470-1485.

Siegenthaler, A., Wangensteen, O. S., Benvenuto, C., Campos, J., & Mariani, S. (2019). DNA metabarcoding unveils multiscale trophic variation in a widespread coastal opportunist. *Molecular Ecology, 28*(2), 232-249.

Smale, D. A., Burrows, M. T., Moore, P., O'Connor, N., & Hawkins, S. J. (2013). Threats and knowledge gaps for ecosystem services provided by kelp forests: a northeast A tlantic perspective. *Ecology and evolution, 3*(11), 4016-4038.

Smith, D. P., & Peay, K. G. (2014). Sequence depth, not PCR replication, improves ecological inference from next generation DNA sequencing. *PloS one, 9*(2), e90234.

Stat, M., Huggett, M. J., Bernasconi, R., DiBattista, J. D., Berry, T. E., Newman, S. J., . . . Bunce, M. (2017). Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. *Scientific Reports, 7*(1), 1-11.

Steneck, R. S., Graham, M. H., Bourque, B. J., Corbett, D., Erlandson, J. M., Estes, J. A., & Tegner, M. J. (2002). Kelp forest ecosystems: biodiversity, stability, resilience and future. *Environmental conservation*, 436-459.

Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. *Molecular Ecology, 21*(8), 1789-1793.

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. doi:10.1111/j.1365-294X.2012.05470.x

Teagle, H., Hawkins, S. J., Moore, P. J., & Smale, D. A. (2017). The role of kelp species as biogenic habitat formers in coastal marine ecosystems. *Journal of Experimental Marine Biology and Ecology, 492*, 81-98.

Thomas, A. C., Tank, S., Nguyen, P. L., Ponce, J., Sinnesael, M., & Goldberg, C. S. (2019). A system for rapid eDNA detection of aquatic invasive species. *Environmental DNA*.

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., . . . Boyer, F. (2016). Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology, 25*(4), 929-942.

van der Loos, L. M., & Nijland, R. (2020). Biases in bulk: DNA metabarcoding of marine communities and the methodology involved. *Molecular Ecology, n/a*(n/a). doi:10.1111/mec.15592

van der Reis, A. L., Laroche, O., Jeffs, A. G., & Lavery, S. D. (2018). Preliminary analysis of New Zealand scampi (Metanephrops challengeri) diet using metabarcoding. *PeerJ, 6*, e5641.

Vergés, A., Steinberg, P. D., Hay, M. E., Poore, A. G., Campbell, A. H., Ballesteros, E., . . . Feary, D. A. (2014). The tropicalization of temperate marine ecosystems: climate-mediated changes in herbivory and community phase shifts. *Proceedings of the Royal Society B: Biological Sciences, 281*(1789), 20140846.

Vlek, H. E., Šporka, F., & Krno, I. j. (2006). Influence of macroinvertebrate sample size on bioassessment of streams. In *The Ecological Status of European Rivers: Evaluation and Intercalibration of Assessment Methods* (pp. 523-542): Springer.

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

Wernberg, T., Bennett, S., Babcock, R. C., De Bettignies, T., Cure, K., Depczynski, M., . . . Hovey, R. K. (2016). Climate-driven regime shift of a temperate marine ecosystem. *Science, 353*(6295), 169-172.

Wernberg, T., Coleman, M. A., Babcock, R. C., Bell, S. Y., Bolton, J. J., Connell, S. D., . . . Shears, N. T. (2019). Biology and Ecology of the Globally Significant Kelp Ecklonia radiata. In *Oceanography and Marine Biology*: Taylor & Francis.

Zhang, G. K., Chain, F. J., Abbott, C. L., & Cristescu, M. E. (2018). Metabarcoding using multiplexed markers increases species detection in complex zooplankton communities. *Evolutionary applications, 11*(10), 1901-1914.

Zhou, J., Wu, L., Deng, Y., Zhi, X., Jiang, Y.-H., Tu, Q., . . . Yang, Y. (2011). Reproducibility and quantitation of amplicon sequencing-based detection. *The ISME journal, 5*(8), 1303-1313.

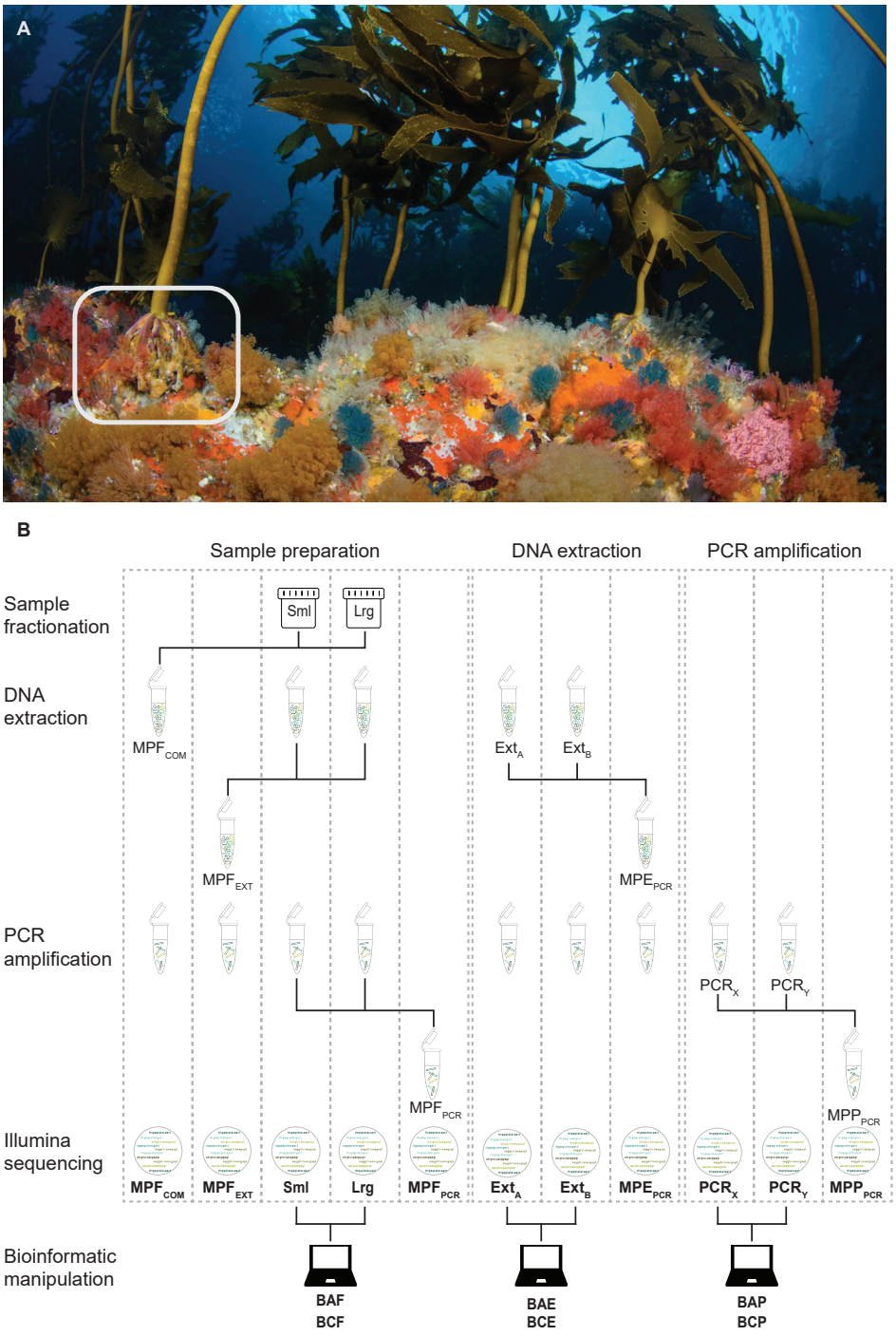
Data accessibility

All FASTQ sequence files are available from the National Center for Biotechnology Information short-read archive database (Bioproject: PRJNA638997, Biosamples: SAMN15220525-SAMN15220620). Associated metadata are also available in the Genomic Obervatories Metadatabase (GEOME; GUID https://n2t.net/ark:/21547/DYo2)

Author contributions

V.A. conducted the laboratory, bioinformatic and biostatistical analysis and wrote the first draft of the manuscript.
J.D.A. secured project funding, collected the samples, designed and performed the biostatistical analysis and edited the manuscript.
L.L. advised the laboratory and bioinformatic protocols, designed the biostatistical analysis and edited the manuscript.
V.A, J.D.A and L.L conceived the experimental design, analyzed the data and designed and prepared figures and tables.

971 **Tables and figures (with captions)**



972

973 Figure 1. A) *Ecklonia radiata* forest and associated hyperdiverse communities making
974 up this ecosystem. The white square highlights the kelp-holdfast which was used as the
975 focal sampling unit in our study. B) Overview of the experimental design to investigate

the impact of alternative decisions and procedures on biodiversity estimates for kelp-holdfast-associated assemblages using community DNA (cDNA) metabarcoding. Experimental treatments were considered during the sample preparation, DNA extraction and PCR amplification steps of the cDNA approach. In bold are the treatments applied to the kelp-holdfast communities which we compared in series of planned contrasts. For each DNA extraction and PCR amplification, two replicates were performed. See Experimental design section for additional details.

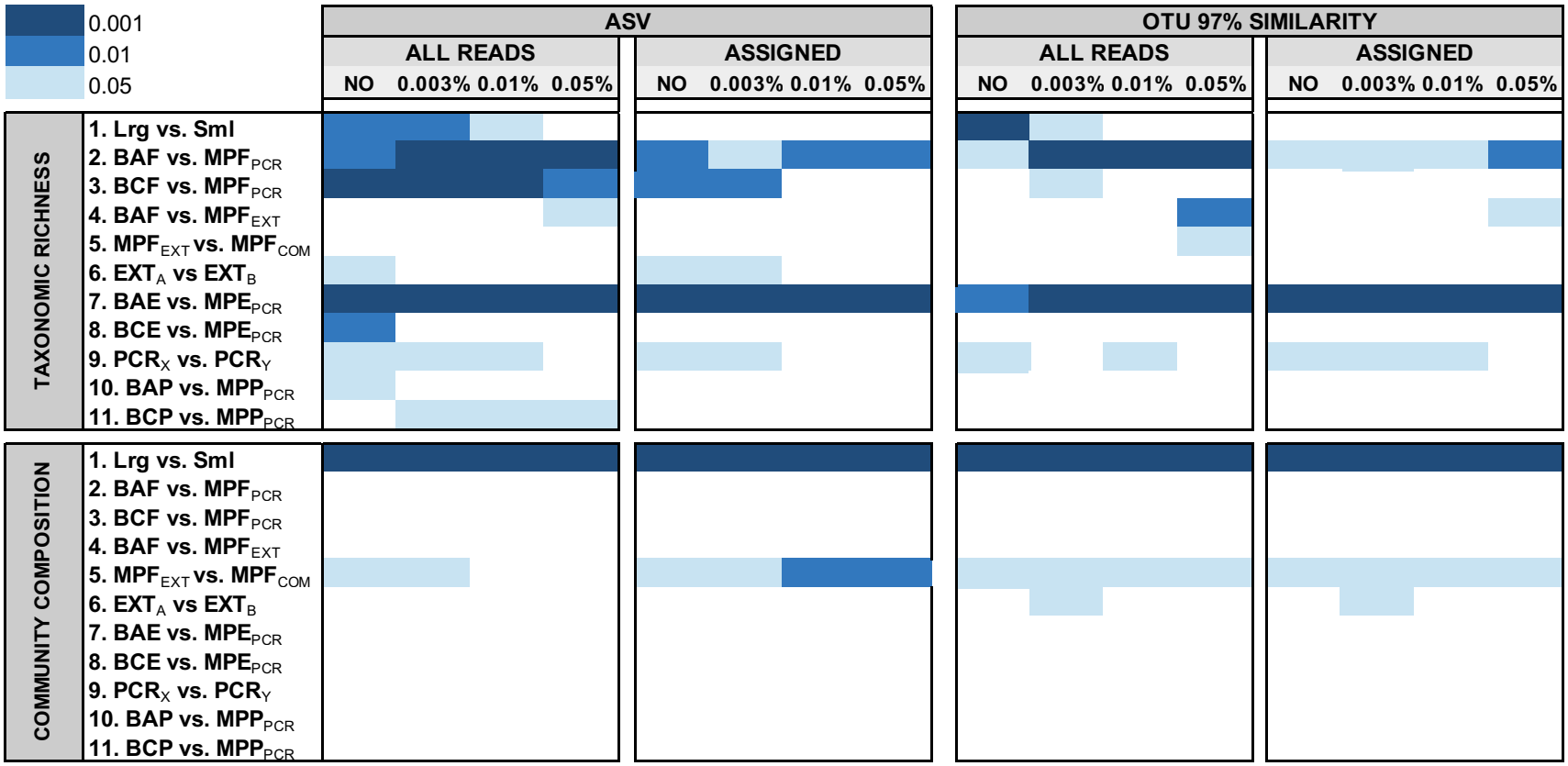


Figure 2. Summary results for the planned contrasts used to examine differences in taxonomic richness (upper table) and community composition (lower table) as a result of different decisions taken during the cDNA approach, including: sample preparation, DNA extraction and PCR amplification. Differences were examined at the ASV- (left table) and OTU-level (right table), considering different filtering thresholds (NO: no filtering, 0.003%, 0.01% and 0.05% minimum read abundance across all samples) and including all reads and only reads taxonomically assigned to Eukaryota. Blue color intensity increases with increasing level of statistical significance (key, top left) and white denotes no statistically significant difference between treatments. A conservative approach was used to find the consensus across three random seeds of each rarefaction performed (see Table S2, Supporting information)

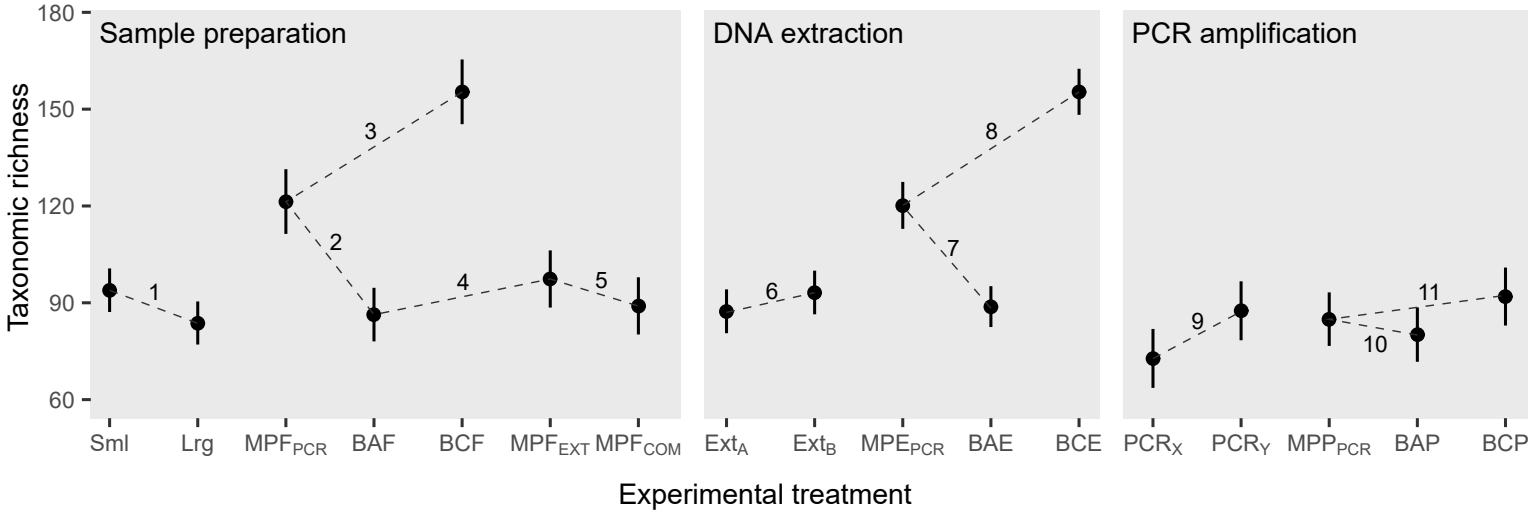
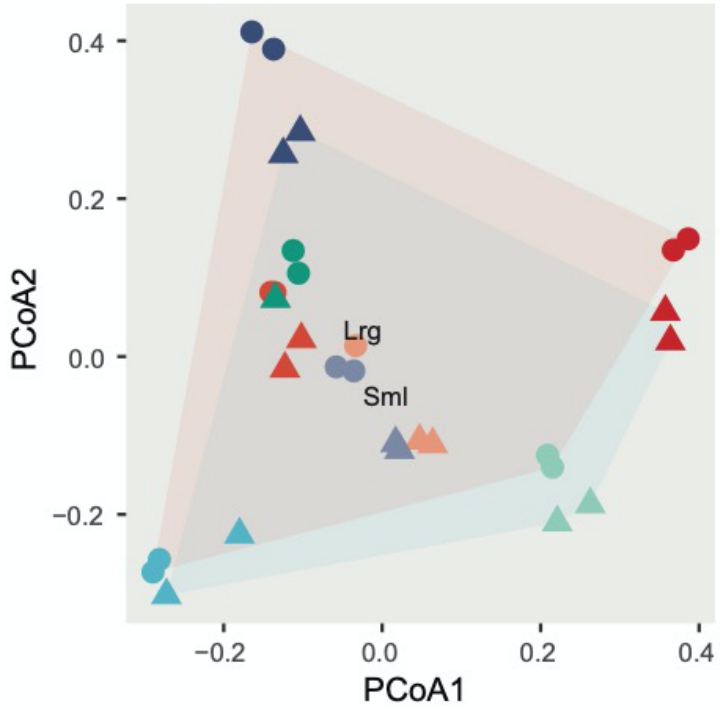


Figure 3. Mean (±SE) taxonomic richness of ASVs filtered at 0.01% minimum read abundance and considering all reads for each of the 11 planned contrasts considered at the sample preparation, DNA extraction and PCR amplification steps of cDNA approach. Dashed lines connect pairs of experimental treatments considered in each numbered contrast (see Experimental design section for details).



999 Figure 4. Principal coordinates of the Jaccard's dissimilarity between the large and
1000 small fractions (shaded polygons) of the sampled kelp-holdfast assemblages. Each
1001 holdfast sample is represented by a different colored symbol. The small size fractions
1002 for each sample are indicated by triangles and the large fractions by circles. The labels
1003 Lrg (large) and Sml (small) indicate the group centroids. Although among sample
1004 differences were large, indicated by the clustering of symbols of the same color within
1005 the PCoA space, differences between size fractions were consistent and significant.

1006

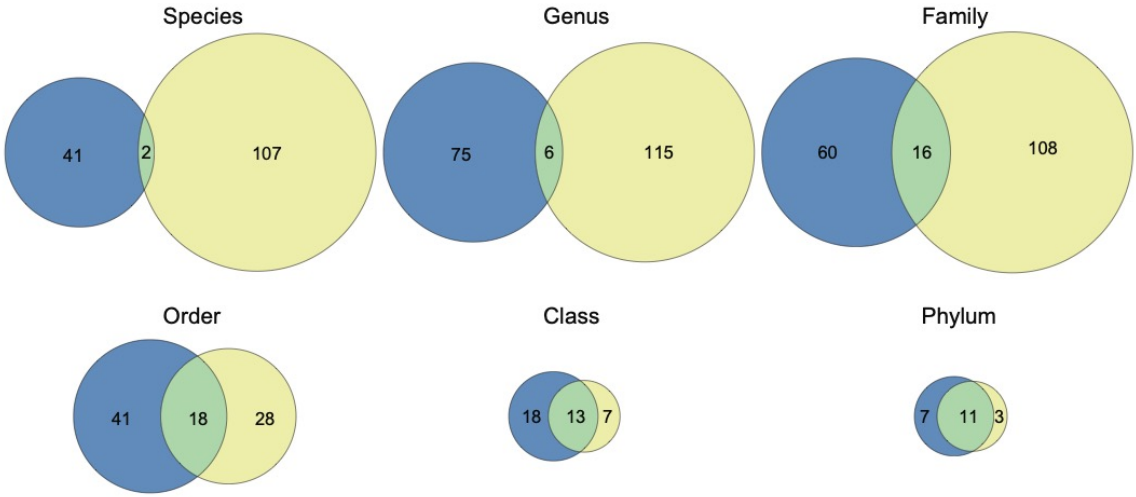


Figure 5. Venn diagrams showing the number of taxa detected by each methodological approach at different taxonomic levels. Results of cDNA metabarcoding are displayed on the left (blue) and traditional morphology-based surveys on the right (yellow) of each Venn diagram, with the number of taxa in common in the intersection (green).

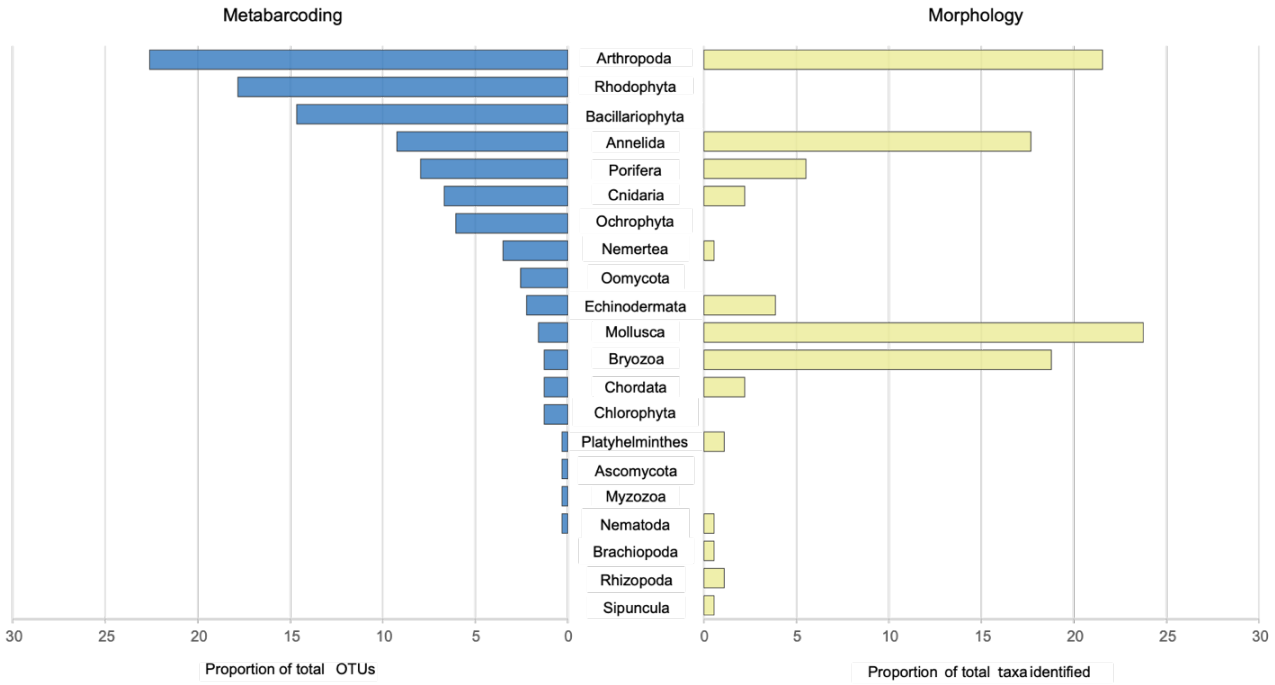


Figure 6. Proportion of taxa identified for each phylum by each methodological approach. DNA metabarcoding on the left (blue) and traditional morphological-based surveys on the right (yellow).