

# Spatial distribution of environmental DNA in a nearshore marine habitat

James L. O'Donnell<sup>\*1</sup>, Ryan P. Kelly<sup>1</sup>, Andrew O. Shelton<sup>2</sup>, Jameal F. Samhouri<sup>3</sup>, Natalie C. Lowell<sup>1,4</sup>, and Gregory D. Williams<sup>5</sup>

<sup>1</sup>School of Marine and Environmental Affairs, University of Washington, 3707 Brooklyn Ave NE, Seattle, Washington 98105, USA

<sup>2</sup>Earth Resource Technology, Inc., Under contract to the Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd E, Seattle, WA 98112, USA

<sup>3</sup>Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd E, Seattle, Washington 98112, USA

<sup>4</sup>School of Aquatic and Fishery Sciences, University of Washington, 1122 NE Boat St, Seattle, Washington 98105, USA

<sup>5</sup>Pacific States Marine Fisheries Commission, Under contract to the Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd E, Seattle, WA 98112, USA

January 13, 2017

## Keywords

metagenomics, metabarcoding, environmental monitoring, molecular ecology, marine, estuarine

---

<sup>\*</sup>jodonnellbio@gmail.com

# Abstract

In the face of increasing threats to biodiversity, the advancement of methods for surveying biological communities is a major priority for ecologists. Recent advances in molecular biological technologies have made it possible to detect and sequence DNA from environmental samples (environmental DNA or eDNA); however, eDNA techniques have not yet seen widespread adoption as a routine method for biological surveillance primarily due to gaps in our understanding of the dynamics of eDNA in space and time. In order to identify the effective spatial scale of this approach in a dynamic marine environment, we collected marine surface water samples from transects ranging from the intertidal zone to 4 kilometers from shore. Using ~~massively parallel sequencing of PCR primers that target a diverse assemblage of metazoans, we amplified a region of mitochondrial 16S amplicons, we identified a diverse community of metazoans and quantified rDNA from the samples and sequenced the products on an Illumina platform in order to detect communities and quantify~~ their spatial patterns using a variety of statistical tools. We find evidence for multiple, discrete eDNA communities in this habitat, and show that these communities decrease in similarity as they become further apart. Offshore communities tend to be richer but less even than those inshore, though diversity was not spatially autocorrelated. Taxon-specific relative abundance coincided with our expectations of spatial distribution in taxa lacking a microscopic, pelagic life-history stage, though most of the taxa detected do not meet these criteria. Finally, we use carefully replicated laboratory procedures to show that laboratory treatments were remarkably similar in most cases, while allowing us to detect a faulty replicate, emphasizing the importance of replication to metabarcoding studies. While there is much work to be done before eDNA techniques can be confidently deployed as a standard method for ecological monitoring, this study serves as a first analysis of diversity at the fine spatial scales relevant to marine ecologists and confirms the promise of eDNA in dynamic environments.

# Introduction

The patterns and causes of variability in ecological communities across space are both seminal and contentious areas of study in ecology (Hubbell, 2001; Anderson et al., 2011). One consistently observed pattern of community spatial heterogeneity is that communities close to one another tend

29 to be more similar than those that are farther apart (Nekola and White, 1999). This decrease  
30 in community similarity with increasing spatial separation is called distance decay and has been  
31 reported from communities of tropical trees (Condit, 2002; Chust et al., 2006), ectomycorrhizal fungi  
32 (Bahram et al., 2013), salt marsh plants (Guo et al., 2015), and microorganisms (Martiny et al.,  
33 2011; Chust et al., 2013; Wetzel et al., 2012; Bell, 2010). Typically, this relationship is assessed by  
34 regressing a measure of community similarity against a measure of spatial separation for a set of  
35 sites at which a set of species' abundances (or presences) is calculated. Yet no existing biodiversity  
36 survey method completely censuses all of the organisms in a given area. The lack of a single 'silver  
37 bullet' method of sampling contributes inconclusiveness to the study of spatial patterning in ecology  
38 (Levin, 1992), and leaves open the possibility of new and more comprehensive methods.

39 From a boat or aircraft, scientists can count whales by sight, but not the krill on which they  
40 feed. For example, towed fishing nets can efficiently sample organisms larger than the mesh and  
41 slower than the boat, but overlook viruses and have undesirable effects on charismatic air-breathing  
42 species. However, DNA-based surveys show great promise as an efficient technique for detecting a  
43 previously unthinkable breadth of organisms from a single sample.

44 Microbiologists have used nucleic acid sequencing to quantify the composition and function of  
45 microbial communities in a wide variety of habitats (Handelsman et al., 1998; Tyson et al., 2004;  
46 Venter et al., 2004; Iverson et al., 2012). To do so, microorganisms are collected in a sample of  
47 environmental medium (e.g. water), their DNA or RNA is isolated and sequenced, and the identity  
48 and abundance of sequences is considered to reflect the community of organisms contained in the  
49 sample, which indirectly estimates the quantity of organisms in an area.

50 Macroorganisms shed DNA-containing cells into the environment (environmental DNA or eDNA)  
51 that can be sampled in the same way (Ficetola et al., 2008; Thomsen et al., 2012). Potentially, eDNA  
52 methods allow a broad swath of macroorganisms to be surveyed from basic environmental samples.  
53 However, the accuracy and reliability of indirect estimates of macroorganismal abundance has been  
54 debated because the entire organisms are not contained within the sample (Cowart et al., 2015).  
55 Concern surrounding eDNA methods is rooted in uncertainty about the attributes of eDNA in the  
56 environment relative to actual organisms (Shelton et al., 2016; Evans et al., 2016). Basic questions  
57 such as how long DNA can persist in that environment and how far DNA can travel remain largely  
58 unknown (but see Klymus et al. (2015); Turner et al. (2015); Strickler et al. (2015); Deiner and

59 Altermatt (2014)) and impede inference about local organismal presence from an environmental  
60 sample. As a result, estimating the spatial and temporal resolution of eDNA studies in the field is  
61 a key step in making these methods practical.

62 The relationship between local organismal abundance and eDNA is further complicated in habi-  
63 tats where the environmental medium itself may transport eDNA away from its source. We know  
64 that genetic material can move away from its source precisely because organisms can be detected  
65 indirectly without being present in the sample (Kelly et al., 2016b). One might reasonably expect  
66 eDNA to travel farther in a highly dynamic fluid such as the open ocean or flowing river than it  
67 would through the sediment at the bottom of a stagnant pond (Deiner and Altermatt, 2014; Shogren  
68 et al., 2016). Yet even studies of extremely dynamic habitats such as coastlines with high wave en-  
69 ergy have found remarkable evidence that eDNA transport is limited enough that DNA methods  
70 can detect differences among communities separated by less than 100 meters (Port et al., 2016).

71 While rigorous laboratory studies have investigated the effects of some environmental factors on  
72 eDNA persistence (Klymus et al., 2015; Barnes et al., 2014; Sassoubre et al., 2016) and the transport  
73 of eDNA in specific contexts (Deiner and Altermatt, 2014), we suggest that field studies comparing  
74 the spatial distribution of communities of eDNA with expectations based on prior knowledge of  
75 organisms' distributions are also critical to developing a working understanding of eDNA in the  
76 real world. Research to date has documented the non-random spatial distribution of meiofaunal  
77 (Fonseca et al., 2014; Guardiola et al., 2016), microbial (Lallias et al., 2015), and extracellular  
78 (Guardiola et al., 2015) eDNA of marine and estuarine sediments, and of microscopic plankton in  
79 open ocean waters (de Vargas et al., 2015). These studies conducted targeted sampling at intermediate  
80 (thousands of meters) to global (thousands of kilometers) scales. Here, we use a grid-based environmental  
81 sampling strategy to assess spatial variability of eDNA in a coastal marine environment at a fine  
82 scale (tens to thousands of meters), using molecular methods that focus on macrobial metazoans.

83 We apply methods derived from community ecology to understand spatial patterns and patchi-  
84 ness of eDNA. The underlying mechanism thought to drive the slope of the distance decay relation-  
85 ship in ecological communities is the rate of movement of individuals among sites, which may be  
86 driven by underlying processes such as habitat suitability. Because eDNA is shed and transported  
87 away from its source, the increased movement of eDNA particles should homogenize community  
88 similarity, and thus erode the distance decay relationship of eDNA communities.

89 Puget Sound is a deep, narrow fjord in Washington, USA, where a narrow band of shallow  
 90 bottom hugs the shoreline and abruptly gives way to a central depth of up to 300 meters. This  
 91 form allows the juxtaposition of communities associated with distinctly different habitats: shallow,  
 92 intertidal benthos, and euphotic pelagic (Burns, 1985). At the upper reaches of the intertidal, the  
 93 shoreline substrate varies from soft, fine sediment to cobble and boulder rubble. Soft intertidal  
 94 sediments are inhabited by burrowing bivalves (Bivalvia), segmented worms (Annelida), and acorn  
 95 worms (Enteropneusta), and in some lower intertidal and high subtidal ranges by eelgrass (*Zostera*  
 96 *marina*) (Kozloff, 1973; Dethier, 2010) . Eelgrass meadows harbor epifaunal and infaunal biota,  
 97 and attract transient species which use the meadows for shelter and to feed on resident organisms.  
 98 Hard intertidal surfaces support a well-documented biota including barnacles (Sessilia) and other  
 99 crustaceans, mussels (Bivalvia:Mytilidae), anemones (~~Actinaria~~Actiniaria), sea stars (Asteroidea),  
 100 urchins (Echinoidea), ~~Bryzeans~~Bryozoans (Ectoprocta), crustaceans (Decapoda), and a variety of  
 101 algae (Dethier, 2010). Hard bottoms of the lower intertidal and high subtidal are home to macroalgae  
 102 such as Laminariales and Desmarestiales which ~~provides~~provide habitat for a distinct community of  
 103 fish and invertebrates. The upper pelagic is home to a diverse assemblage of microscopic plankton  
 104 including diatoms, copepods, and larvae (Strickland, 1983), as well as transitory fish and marine  
 105 mammals.

106 We took advantage of this setting to explore the spatial variation and distribution of marine  
 107 eDNA communities. Using PCR-based methods and massively parallel sequencing, we surveyed  
 108 mitochondrial 16S sequences from a suite of marine animals in water samples collected over a grid  
 109 of sites extending from the shoreline out to 4 kilometers offshore in Puget Sound, Washington, USA.  
 110 We leverage this sampling design to perform ~~the first~~an explicitly spatial analysis of eDNA-derived  
 111 community similarity. We investigate two primary objectives. First we examine the spatial pattern-  
 112 ing of eDNA and determine the degree to which eDNA community similarity can be predicted by  
 113 physical proximity. We expect that physical proximity will be a strong predictor of community sim-  
 114 ilarity, and that community differences can be detected over small distances. Second, we examine  
 115 the distribution of diversity from eDNA data, and compare it to our expectations based on distri-  
 116 butions of macrobial communities. We expect that distinct eDNA communities exist in this setting,  
 117 and that their spatial distribution coincides with that of adult macrobial organisms. Because of the  
 118 vastly different communities of benthic macrobial metazoans as a function of distance from shore,

119 we expect that more than one eDNA community is present across our 4 kilometer sampling grid,  
120 and that communities change as a function of distance from shore. For this reason, we examine two  
121 diversity measures of eDNA communities that have been widely used to reveal broad scale patterns  
122 based on macrobiota in many ecological systems. Finally, we identify the taxa represented in the  
123 eDNA communities, which span a range of life-history characteristics, and we expect that the spatial  
124 distribution of eDNA will most closely resemble the distribution of adults in taxa with low dispersal  
125 potential.

## 126 **Methods**

127 There are seven discrete steps to our methodology: (1) Environmental sample collection, (2) isolation  
128 of particulates from water via filtration, (3) isolation of DNA from filter membrane, (4) amplification  
129 of target locus via PCR, (5) sequencing of amplicons, (6) bioinformatic translation of raw sequence  
130 data into tables of sequence abundance among samples, and (7) community ecological analyses of  
131 eDNA. We provide brief overviews of these steps here, and encourage the reader to review the fully  
132 detailed methods presented in the supplementary material (Supplemental Material).

### 133 **Environmental Sampling**

134 Starting from lower-intertidal patches of *Zostera marina*, we collected water samples at 1 meter  
135 depth from 8 points (0, 75, 125, 250, 500, 1000, 2000, and 4000 meters) along three parallel transects  
136 separated by 1000 meters (24 sample locations total; Figure 1). Samples were collected by attaching  
137 bottles to a PVC pole and lowering it over the side of a boat over the span of one hour on 27 June  
138 2014. To destroy residual DNA on equipment used for field sampling and filtration, we washed  
139 with a 1:10 solution of household bleach (8.25% sodium hypochlorite; 7.25% available chlorine) and  
140 deionized water, followed by thorough rinsing with deionized water. Each environmental sample  
141 was collected in a clean 1 liter high-density polyethylene bottle, the opening of which was covered  
142 with 500 micrometer nylon mesh to prevent entry of larger particles. Immediately after collecting  
143 the sample, the mesh was replaced with a clean lid and the sample was held on ice until filtering.

## 144 Filtration

145 One liter from each water sample was filtered in the lab on a clean polysulfone vacuum filter  
146 holder fitted with a 47 millimeter diameter cellulose acetate membrane with 0.45 micrometer pores.  
147 Filter membranes were moved into 900 microliters of Longmire buffer (Longmire et al., 1997) using  
148 clean forceps and stored at room temperature ~~(?)~~[\(Renshaw et al., 2015\)](#). To test for the extent of  
149 contamination attributable to laboratory procedures, we filtered three replicate 1 liter samples of  
150 deionized water. These samples were treated identically to the environmental samples throughout  
151 the remaining protocols.

## 152 DNA Purification

153 DNA was purified from the membrane following a phenol:chloroform:isoamyl alcohol protocol follow-  
154 ing Renshaw ~~(?)~~[\(Renshaw et al., 2015\)](#). Preserved membranes were incubated at ~~65°C~~[65 °C](#) for 30  
155 minutes before adding 900 microliters of phenol:chloroform:isoamyl alcohol and shaking vigorously  
156 for 60 seconds. We conducted two consecutive chloroform washes by centrifuging at 14,000 rpm for 5  
157 minutes, transferring the aqueous layer to 700 microliters chloroform, and shaking vigorously for 60  
158 seconds. After a third centrifugation, 500 microliters of the aqueous layer was transferred to tubes  
159 containing 20 microliters 5 molar NaCl and 500 microliters 100% isopropanol, and frozen at ~~-20°C~~  
160 [-20 °C](#) for approximately 15 hours. Finally, all liquid was removed by centrifuging at 14000 rpm for  
161 10 minutes, pouring off or pipetting out any remaining liquid, and drying in a vacuum centrifuge  
162 at ~~45°C~~[45 °C](#) for 15 minutes. DNA was resuspended in 200 microliters of ultrapure water. Four  
163 replicates of genomic DNA extracted from tissue of a species absent from the sampled environment  
164 (*Oreochromis niloticus*) served as positive control for the remaining protocols.

## 165 PCR Amplification

166 ~~From each DNA sample, we amplified~~[We chose a primer set that amplifies](#) an approximately  
167 115 base pair (bp) region of the mitochondrial ~~gene encoding~~ 16S ~~RNA using a~~[rRNA gene in at](#)  
168 [least 10 metazoan phyla from this habitat, excludes non-metazoans, and resolves taxonomy to the](#)  
169 [family level in most cases using a public sequence database \(Kelly et al., 2016a\)](#). We used a two-  
170 step polymerase chain reaction (PCR) protocol described by O'Donnell et al. (2016) [to generate 4](#)

171 [replicate products from each DNA sample](#). In the first set of reactions, primers were identical in ev-  
172 ery reaction (forward: AGTTACYYTAGGGATAACAGCG; reverse: CCGGTCTGAACTCAGAT-  
173 CAYGT); primers in the second set of reactions included these same sequences but with 3 variable  
174 nucleotides (NNN) and an index sequence on the 5' end (see [Sequencing Metadata](#)). We used the  
175 program OligoTag (Coissac, 2012) to generate 30 unique 6-nucleotide index sequences differing by  
176 a minimum Hamming distance of 3 (see [Sequencing Metadata](#)). Indexed primers were assigned to  
177 samples randomly, with the identical index sequence on the forward and reverse primer to avoid  
178 errors associated with dual-indexed multiplexing (Schnell et al., 2015). In a UV-sterilized hood,  
179 we prepared 25 microliter reactions containing 18.375 microliters ultrapure water, 2.5 microliters  
180 10x buffer, 0.625 microliters deoxynucleotide solution (8 millimolar), 1 microliter each forward and  
181 reverse primer (10 micromolar, obtained lyophilized from Integrated DNA Technologies (Coralville,  
182 IA, USA)), 0.25 microliters Qiagen HotStar Taq polymerase, and 1.25 microliter genomic [or](#) eDNA  
183 template at 1:100 dilution in ultrapure water. PCR thermal profiles began with an initialization  
184 step (~~95~~[95](#) °C; 15 min) followed by cycles (40 and 20 for the first and second reaction, respectively)  
185 of denaturation (~~95~~[95](#) °C; 15 sec), annealing (~~61~~[61](#) °C; 30 sec), and extension (~~72~~[72](#) °C; 30 sec).  
186 20 identical PCRs were conducted from each DNA extract using non-indexed primers; these were  
187 pooled into 4 groups of 5 in order to ensure ample template for the subsequent PCR with indexed  
188 primers. In order to isolate the fragment of interest from primer dimer and other spurious frag-  
189 ments generated in the first PCR, we used the AxyPrep Mag FragmentSelect-I kit with solid-phase  
190 reversible immobilization (SPRI) paramagnetic beads at 2.5x the volume of PCR product (Axygen  
191 BioSciences, Corning, NY, USA). A 1:5 dilution in ultrapure water of the product was used as tem-  
192 plate for the second reaction. PCR products of the second reaction were purified using the Qiagen  
193 MinElute PCR Purification Kit (Qiagen, Hilden, Germany). Ultrapure water was used in place of  
194 template DNA and run along with each batch of PCRs to serve as a negative control for PCR; none  
195 of these produced visible bands on an agarose gel. In total, four separate replicates from each of  
196 31 DNA samples were carried through the two-step PCR process for a total of 124 sequenced PCR  
197 products. These were combined with additional samples from other projects, totaling 345 samples  
198 for sequencing.



## 199 DNA Sequencing

200 Up to 30 PCR products were combined according to their primer index in equal concentration into  
201 one of 14 pools, and 150 nanograms from each were prepared for library sequencing using the KAPA  
202 high-throughput library prep kit with real-time library amplification protocol (KAPA Biosystems,  
203 Wilmington, MA, USA). Each of these ligated sequencing adapters included an additional 6 base  
204 pair index sequence (NEXTflex DNA barcodes; BIOO Scientific, Austin, TX, USA). Thus, each  
205 PCR product was identifiable via its unique combination of index sequences in the sequencing  
206 adapters and primers. Fragment size distribution and concentration of each library was quantified  
207 using an Agilent 2100 BioAnalyzer. Libraries were pooled in equal concentrations and sequenced  
208 for 150 base pairs in both directions (PE150) using an Illumina NextSeq at the Stanford Functional  
209 Genomics Facility(~~machine NS500615, run 115, flowcell H3LFLAFX~~), where 20% PhiX Control  
210 v3 was added to act as a sequencing control and to enhance sequencing depth [by increasing sequence](#)  
211 [diversity](#). Raw sequence data in fastq format is publicly available (see Data Availability).

## 212 Sequence Data Processing (Bioinformatics)

213 Detailed bioinformatic methods are provided in the supplemental material, and analysis scripts  
214 used from raw sequencer output onward can be found in the public project directory (see [Analysis](#)  
215 [Scripts](#)). Briefly, we performed five steps to process the sequence data: (1) Merge paired-end  
216 reads, (2) eliminate low-quality reads, (3) eliminate PCR artifacts (chimeras), (4) cluster reads by  
217 similarity into operational taxonomic units (OTUs), and (5) match observed sequences to taxon  
218 names. Additionally, we checked for consistency among PCR replicates, excluded extremely rare  
219 sequences, and rescaled (rarefied) the data to account for differences in sequencing depth. The data  
220 for input to further analyses are a contingency table of the mean count of unique sequences, OTUs,  
221 or taxa present in each environmental sample.

## 222 Ecological Analyses

223 After gathering the data, we use the eDNA community observed at each location to make inferences  
224 about the spatial patterning of eDNA communities. We use statistical tools from community ecology  
225 to assess the spatial structure of eDNA communities. We report similarity (1- dissimilarity) rather

226 than dissimilarity in all cases for ease of interpretation.

## 227 **Objective 1: Community similarity as a function of distance**

### 228 **Distance Decay**

229 To address our first objective and determine whether or not nearby samples are more similar than  
230 distant ones, we fit a nonlinear model to represent decreasing community similarity with distance.  
231 We calculated the pairwise Bray-Curtis similarity (1 - Bray-Curtis dissimilarity) between eDNA  
232 communities using the R package *vegan* (Oksanen et al., 2016) and the great circle distance between  
233 sampling points using the Haversine method as implemented by the R package *geosphere* (Hijmans,  
234 2016). This model is similar to the Michaelis-Menten function, but with an asymptote fixed at 0:

$$y_{ij} = \frac{AB}{B + x_{ij}} \quad (1)$$

235 Where the relationship between community similarity ( $y_{ij}$ ) and spatial distance ( $x_{ij}$ ) between  
236 observations  $i$  and  $j$  is determined by the similarity of samples at distance 0 ( $A$ ), and the distance at  
237 which half the total change in similarity is achieved ( $B$ ). This allows for samples collected very close  
238 together (near 0) to have similarity significantly less than one. We assessed model fit using the R  
239 function *nls* (R Core Team, 2016), using the *nl2sol* algorithm from the *Port* library to solve separable  
240 nonlinear least squares using analytically computed derivatives (<http://netlib.org/port/nsg.f>). We  
241 set bounds of 0 and 1 for the intercept parameter and a lower bound of 0 for the distance at half  
242 similarity; starting values of these parameters were 0.5 and  $x_{max}/2$ , respectively. We calculated  
243 a 95% confidence interval for the parameters and the predicted values using a first-order Taylor  
244 expansion approach implemented by the function *predictNLS* in the R package *propagate* (Spiess,  
245 2014).

246 There are other conceptually reasonable forms to expect the space-by-similarity relationship  
247 to take; we present these in the supplemental material along with alternative data subsets and  
248 similarity indices (see Supplemental Material).

## 249 **Objective 2: Spatial distribution of diversity**

### 250 **Community Classification**

251 To determine the spatial distribution and variation of eDNA communities (objective 2), we used  
252 multivariate classification algorithms. We simultaneously assessed the existence of distinct com-  
253 munity types and the membership of samples to those community types using an unsupervised  
254 classification algorithm known as partitioning around medoids (PAM; sometimes referred to as k-  
255 medoids clustering) (Kaufman and Rousseeuw, 1990), as implemented in the R package cluster  
256 (Maechler et al., 2016). The classification of samples to communities was made on the basis of  
257 their pairwise Bray-Curtis similarity, calculated using the function `vegdist` in the R package `vegan`  
258 (Oksanen et al., 2016). Other distance metrics were evaluated but had no appreciable effect on the  
259 outcome of the analysis (Figure 8). In order to chose an optimal number of clusters ( $K$ ), we evalu-  
260 ated the distribution of silhouette widths, a measure of the similarity between each sample and its  
261 cluster compared to its similarity to other clusters. We repeated the analysis using fuzzy clustering  
262 (FANNY, (Kaufman and Rousseeuw, 1990); however, the results were qualitatively similar to the  
263 results using PAM so we omit them here.

### 264 **Aggregate Measures of Diversity**

265 We calculated two measures of diversity, richness and evenness, to ask if aggregate metrics of the  
266 eDNA community showed evidence of spatial patterning. Richness is a measure of the number of  
267 distinct types of organisms present and so ranges from 1 (only one taxon observed) to  $S$ , the number  
268 of taxa observed across all samples. To calculate the evenness of the distribution of abundance of  
269 taxa in a sample, we used the complement of the Simpson (1949) index ( $1 - \sum p_i^2$ , where  $p_i$  is the  
270 proportional abundance of taxon  $i$ ). The values of this index ranges from 0 to 1, with the value  
271 interpreted as the probability that two sequences randomly selected from the sample will belong to  
272 different taxa; thus, larger values of the index indicate more evenly divided communities (Magurran,  
273 2003). We calculated Moran's I for both diversity metrics to test for spatial autocorrelation. We  
274 also tested for a linear effect of log-transformed distance from shore on each measure of diversity to  
275 ask how diversity changes over this strong environmental gradient.

## 276 Taxon and Life History Patterns

277 After assigning taxon names to the abundance data, we plotted the distribution in space of a  
278 selection of taxa to compare with our expectations on the basis of adult distributions (objective 2).  
279 Our aim was to understand where each taxon occurred in the greatest proportional abundance, and  
280 its distribution in space relative to that maximum. Thus, we rescaled each sample to proportional  
281 abundance, extracted the data from a single taxon, and scaled those values between 0 and 1. We  
282 collated life history characteristics for each of the major taxonomic groups recovered, including  
283 dispersal range of the gametes, larvae, and adults, adult habitat type and selectivity, and adult  
284 body size. ~~Dispersal range was given as~~ For each life history stage of each taxon group, we made an  
285 order-of-magnitude approximation of the scale of dispersal: ~~for~~. For example, internally fertilized  
286 species were assigned a gamete range of 0 km, while broadcast spawners were assigned a gamete  
287 range of 10 km. Similarly, adult range size was approximated as 0 km (sessile), 1 km (motile but  
288 not pelagic), or 10 km (highly mobile, pelagic). Variables were specified as 'multiple' for ~~groups~~ life  
289 history stages known to span more than 1 magnitude of range size. For groups to which sequences  
290 were annotated with high confidence, but for which life history strategy is diverse or poorly known  
291 (e.g. families in the phylum Nemertea), we used conservative, coarse approximations at a higher  
292 taxonomic rank (see Life History Data). These data were used to contextualize group-specific spatial  
293 distributions and inform expectations based on known adult distributions.

## 294 Results

### 295 Sequence Data Processing (Bioinformatics)

296 Preliminary sequence analysis strongly suggested that the observed variation among environmental  
297 samples reflects true variation in the environment, rather than variability due to lab protocols, for  
298 the following reasons (note that all value ranges are reported as mean plus and minus one standard  
299 deviation). First, all libraries passed the FastQC per-base sequence quality filter, generating a total  
300 of 371,576,190 reads passing filter generated in each direction. Second, samples in this study were  
301 represented by an adequate number of reads ( $333,537.9 \pm 112,200.5$ ), with no individual sample  
302 receiving fewer than 130,402 reads. Third, there was a very low frequency of cross-contamination

303 from other libraries into those reported here ( $5e-05 \pm 8e-05$ ; max proportion 0.00034). Fourth, after  
304 scaling all samples to the same sequencing depth, OTUs with abundance greater than 178 reads  
305 (0.14% of a sample's reads) experienced no turnover among PCR replicates within a sample. Fifth,  
306 sequence abundances among PCR replicates within water samples were remarkably consistent. A  
307 single sample had low similarity among PCR replicates (0.659) after removing this outlier, the  
308 lowest mean similarity among replicates within a sample was 0.966. Overall similarities among  
309 PCR replicates within a sample were extremely high ( $0.976 \pm 0.013$ ), and far higher than ~~that of~~  
310 ~~than those~~ among samples ( $0.3 \pm 0.16$ ). Across PCR replicates, each sample was represented by at  
311 least 781425 reads in the raw data and contained between 111 and 443 rarefied OTUs (Supplemental  
312 Figure 10).

## 313 Ecological Analyses

### 314 Distance Decay

315 Physical proximity is a good predictor of eDNA community similarity: Similarity decreased from  
316 0.40 (95%CI = 0.36, 0.45) to half that amount at 4500 meters (95%CI = 2900, 7500) (Figure 2).

### 317 Community Classification

318 Despite a clear trend in community similarity as a function of spatial separation, the results from  
319 our classification analysis are difficult to interpret. The silhouette analysis indicated the presence  
320 of 8 distinct communities; however, the gain in mean silhouette width from 2 was small (0.1), and  
321 lacked a distinctive peak (Figure 4), indicating substantial uncertainty in the clustering algorithm.  
322 Thus, we present the results of cluster assignment for both  $K = 2$  and  $K = 8$  to illustrate the  
323 range of results (Figure 3). Excluding taxa which occur in only one site had no discernible effect  
324 on the outcome of the PAM analysis (number of clusters, assignment to clusters). While there was  
325 no distinct spatial divide indicating the presence of an inshore versus an offshore community, one  
326 of the two communities (at  $K = 2$ ) occurred in only 2 out of 18 samples inside 1000 meters from  
327 shore, and never occurred within 125 meters of shore, suggesting the presence of an inshore and  
328 offshore community.

## 329 Diversity in Space

330 Sites offshore tend to be less rich and more even than those inshore (Figure 6). Mean OTU richness  
331 declined by 1.42 per 1000 meters from a mean of 17.6 taxa (95%CI = 2.15) inshore to 11.9 taxa  
332 (95%CI = 4.31) at offshore locations ( $p = 0.0415$ ; Figure 6). Evenness ~~the probability that two~~  
333 ~~reads chosen at random from a sample belong to different species,~~ increased by .0666 per 1000  
334 meters from 0.225 (95%CI = 0.0558) to 0.491 (95%CI =  $\pm 0.112$ ), indicating that sequence reads  
335 were less evenly distributed among taxa in offshore samples ( $p \ll 0.05$ ; Figure 6). There was no  
336 evidence for spatial autocorrelation for any of the diversity metrics (Moran's I,  $p > 0.05$ ; Figure 5).

## 337 Taxon and Life History Patterns

338 We were able to assign a taxon name with confidence to 136 of 146 OTU sequences. The vast ma-  
339 jority of sequences (97.6%) and OTUs (96.9%) were matched to organisms that have high potential  
340 for dispersal at either the gamete, larval, or adult stage, making it impossible to determine whether  
341 the source of that DNA was adults with well-documented spatial patterns (e.g. sessile nearshore  
342 specialists) or highly mobile early life history stages. Of the 6 OTUs for which dispersal is limited  
343 during all life history stages, only 2 occurred in more than two samples, precluding a quantita-  
344 tive comparison of spatial dispersion based on life history characteristics. These were assigned to  
345 *Cymatogaster aggregata*, a viviparous nearshore fish with internal fertilization, and *Cupolaconcha*  
346 *meroclista*, a sessile Vermetid gastropod with presumed internal fertilization and short larval dis-  
347 persal (Strathmann and Strathmann, 2006; Phillips and Shima, 2010; Calvo and Templado, 2004).  
348 *Cymatogaster aggregata* was distinctly more abundant close to shore, with no sequences occurring  
349 in any sample beyond 250 meters (Figure 7). *Cupolaconcha meroclista* showed no such distinct  
350 spatial trend, occurring in nearly equal abundance at three sites, 75, 500, and 2000 meters from  
351 shore. An additional species that was highly abundant in the sequence data, the krill *Thysanoessa*  
352 *raschii*, has pelagic adults, highly seasonal reproduction, and sinking eggs; their distribution was  
353 consistent with our expectations based on a tendency of adults to aggregate offshore. Finally, the  
354 two most abundant taxa in the dataset were the mussel genus *Mytilus* and the Barnacle order Ses-  
355 silia; the adults of both taxa are sessile and occur exclusively on hard intertidal substrata but have  
356 highly motile larvae. Because large-scale dispersal could not be ruled out for the vast majority of

~~taxa, subsetting the community data by taxonomic group had no qualitative effect on the spatial  
patterning or diversity metrics, and we omit those results here.~~

## Discussion

Indirect surveys of organismal presence are a key development in ecosystem monitoring in the face of increased anthropogenic pressure and dwindling resources for ecological research. Monitoring of organisms using environmental DNA is an especially promising method, given the rapid pace of advancement in technological innovation and cost efficiency in the field of DNA sequencing and quantification. ~~For the first time in a marine environment, we~~ We document four key patterns: (1) eDNA communities far from one another tend to be less similar than those that are nearby, (2) distinct eDNA communities exist and are distributed in a non-random fashion, (3) diversity declines with distance from shore, and (4) spatial patterning of eDNA is associated with taxon-specific life history characteristics.

### **(1) Communities far from one another tend to be less similar than those that are nearby**

We demonstrate that ~~more~~-distant locations have ~~less similar~~ less similar eDNA communities than ~~more~~-proximate locations in Puget Sound, a dynamic marine environment. Our finding is in line with observations based on traditional surveys of terrestrial plants and fungi (Nekola and White, 1999; Bahram et al., 2013; Condit, 2002; Chust et al., 2006) and of microorganisms in freshwater (Wetzel et al., 2012), marine (Chust et al., 2013), and estuarine (Martiny et al., 2011) environments. To our knowledge, it is the first to report such a pattern using massively parallel sequencing of environmental DNA in the marine environment, and the first using any technique to describe this pattern from microbial metazoans. We note that the theoretical expectation is that samples at very close distance be nearly completely similar, while our samples separated by the 50 meters were only 40% similar. We interpret this to reflect the highly dynamic nature of this environment, which could cause DNA to be distributed quickly from its source, eroding the rise in similarity at small distances. At the same time, community similarity decreased to very low levels at larger scales, indicating that DNA distribution is not completely unpredictable. This finding implies that the effectively sampled

area of individual water samples for eDNA analysis is likely to be quite small ( $<100\text{m}$ ) in this nearshore environment. Our estimated distance-decay relationship does indicate that proximate samples are more similar than distant samples, but we suggest this pattern is partially obscured by other factors, including signal from mobile, microscopic life-stages.

## (2) Distinct eDNA communities exist and are distributed in a non-random fashion

We demonstrate strong evidence for distinct community types and the non-random spatial patterning of those communities. While the spatial distributions of communities is surprising if one were concerned only with the macroscopic life stages of metazoans, it indeed does align with the broader view that even offshore pelagic communities are comprised of and influenced by nearshore organisms. This result underscores the idea that areas immediately offshore act as ecotones, a mixing zone of taxa characteristic of benthic and pelagic environments. While there was no distinct break in community types between onshore and offshore sites, there was some clustering of community types that may be explained by oceanographic features such as nearshore eddies generated by strong tidal exchange in a steep bathymetric setting (Yang and Khangaonkar, 2010). It would be useful to better understand such features during the period of sampling, by way of oceanographic monitoring devices. Finally, the uncertainty in identification of the number of distinct clusters to best characterize the community underlines the difficulty of identifying community patterns with the number of taxonomic groups considered here. We suspect that the signature of eDNA from microscopic life-stages may explain our inability to easily detect spatial community level patterns that align with our initial expectations.

## (3) Richness declines and evenness increases with distance from shore

We ~~detected a general pattern of declining richness and increasing evenness with increasing distance offshore~~ found that richness declined while evenness increased with distance from shore. Such a pattern is consistent with many other ecosystems which show strong clines in diversity metrics over environmental gradients. ~~The coastal ocean is a highly productive and diverse ecosystem (Ray, 1988).~~ However, our study is novel in that it corroborates a cline well-known on macroscales for macrobiota on a much smaller spatial scale for microscopic animals, suggesting that there may be a self-similarity across scales in diversity patterning (Levin, 1992). The coastal ocean is a highly



productive and diverse ecosystem, where biomass is concentrated most heavily along the bottom and shoreline (Ray, 1988). This differential in biomass concentration from the shoreline to open waters may contribute to the opposing trends we detected. Where particles (organisms, tissues, and cells) are sparse, fewer would be collected per sample of constant volume, thus decreasing the probability of drawing as many types (richness) and increasing the probability that any two particles originate from the same type (evenness). Intriguingly, the cline in diversity from inshore to offshore was not determined by shared changes in communities as one moved offshore; the classification analysis suggested a fair amount of differences among communities at a given offshore distance (Figure 3). ~~Furthermore, the uncertainty in identification of the number of distinct clusters to best characterize the community underlines the difficulty of identifying community patterns with the number of taxonomic groups considered here. We suspect that the signature of eDNA from microscopic life-stages may explain our inability to easily detect spatial community level patterns that align with our initial expectations.~~

#### (4) Spatial patterning of eDNA is associated with taxon-specific life history ~~characteristics.~~

In contrast to our expectations, other taxa including species with sessile adult stages restricted to benthic hard substrates (e.g. barnacles, mussels) are among the most abundant taxa at sites furthest from shore. However, the larvae and gametes of these taxa are abundant, pelagic, and can be transported long distances by water movement (Strathmann, 1987). This indicates that we likely detected DNA of their pelagic phase gametes and larvae. It is always possible that DNA of adults was advected over long distances and detected offshore but in light of our results with krill and surfperch, we view this as unlikely. We interpret our results as evidence that the chaotic spatial distribution of eDNA communities (Figure 3) results from our primers' affinity for many species which at some point exist as microscopic pelagic gametes or larvae. Our results emphasize that expected results based on easily visually observed individuals or detectable with traditional sampling gear such as nets may be very different from results using eDNA. This does caution that eDNA surveys may have different purposes and may not be directly comparable to existing surveys (Shelton et al., 2016).

We acknowledge that sampling artifacts may have affected our results. For example if entire multicellular individuals were captured in our samples, their DNA could be in much greater density

441 than eDNA, affecting the observed community. Our sampling bottles excluded particles larger than  
442 500 micrometers, but gametes and very small larvae could have gained entry. It is possible that  
443 even a single small individual, containing many thousand mitochondria, would overwhelm the signal  
444 of another species from which hundreds of cells had been sloughed from many, larger individuals.  
445 Data on larval size distribution at the time of sampling from each species in our data set would  
446 allow us to estimate the frequency of such events. Nevertheless, it is precisely the sensitivity to  
447 small particles that makes the eDNA approach powerful, so we are reluctant to recommend that  
448 aquatic eDNA sampling use finer pre-filtering. Instead, we emphasize the importance of designing  
449 and selecting primer sets that selectively amplify target organisms. In the case of the present study,  
450 in order to recover patterns matching our expectations, this would be non-transient, benthic marine  
451 organisms lacking any pelagic life stage.

452 The marker we chose for this study detects a wide variety of metazoans while excluding other  
453 more common taxa; however, it does not effectively discriminate among species within a higher group  
454 in all cases. Other markers, such as mitochondrial cytochrome c oxidase subunit 1 (COX1, CO1,  
455 or COI) may provide adequate species-level resolution in some metazoan groups, but have other  
456 shortcomings including taxon dropout (Deagle et al., 2014) and amplification of more abundant  
457 non-metazoans, as we discovered in an accompanying study (Kelly et al., 2016a). Both have undesirable  
458 effects of biasing estimates of diversity. In our case, it is possible that the lumping of multiple  
459 species into one group underestimates the true richness of the group and of the entire sample, in  
460 turn obscuring true underlying patterns of diversity. In the case of COX1, well-documented primer  
461 biases cause failure to amplify some taxa, particularly in mixed samples, with the same result  
462 (Deagle et al., 2014). In fact, even surveys relying on traditional capture techniques (e.g. seine  
463 nets) and morphological characteristics are subject to biases imposed by the sampling gear (e.g.  
464 mesh size), the observer (e.g. taxonomic expertise), and organisms (e.g. morphologically cryptic  
465 species). Similarly, no single molecular marker adequately and effectively samples all taxa without  
466 bias (Drummond et al., 2015), and thus the choice of marker is an important and context-dependent  
467 one. Until whole-genome sequencing of individual cells is a reality, the tradeoffs between taxonomic  
468 breadth and resolution will continue to be problematic for metabarcoding studies, just as they are  
469 for more traditional ecological survey methods (Kelly et al., 2016a).

470 Our results also highlight the need for curated life-history databases. As technological advances

471 increase the speed and throughput of DNA sequencing and sequence processing, making sense of  
472 these data in a timely manner requires that natural history data be stored in standard formats in  
473 centralized repositories. The rate at which we can make sense of high-throughput survey methods  
474 will be limited by our ability to collate auxiliary data. Databases such as Global Biodiversity  
475 Information Facility (GBIF), Encyclopedia of Life (EOL), and FishBase (Parr et al., 2014; Froese  
476 and Pauly, 2016) contain records of taxonomy, occurrence, and other rudimentary data types, but  
477 there is no centralized, standardized repository for even basic natural history data such as body  
478 size. As NCBI's nucleotide and protein sequence database (GenBank) has facilitated transformative  
479 studies in diverse fields, an ecological analog would be a boon for biodiversity science.

480 Surveys based on eDNA are intensely scrutinized because of the danger that the final data are  
481 subject to complicated laboratory and bioinformatic procedures. Finding virtually no variability  
482 among lab and bioinformatic treatments from the point of PCR onward, we were confident our  
483 results represented actual field-based differences among samples. However, we note that one PCR  
484 replicate had a clear signal of contamination in that the sequence community was extremely similar  
485 to those from a different environmental sample. The source of this error is difficult to identify, but  
486 seems most likely to be an error during PCR preparation, either in assignment or pipetting during  
487 preparation of indexed primers. While the remainder of our results would be largely unchanged  
488 had we sequenced a single replicate per environmental sample, we believe the sequencing of PCR  
489 replicates is critical for ensuring data quality in eDNA sequencing studies.

490 While there is much work to be done before eDNA techniques can be confidently deployed as a  
491 standard method for ecological monitoring, this study serves as a first analysis of diversity at the  
492 fine spatial scales that are likely to be relevant to eDNA work in the field across a range of study  
493 systems.

## 494 Acknowledgements

495 We wish to thank [Linda Park](#), Robert Morris, E. Virginia Armbrust, and James Kralj. [The](#)  
496 [manuscript was improved by suggestions from editor Magnus Johnson, and reviewers Owen Wangenstein](#)  
497 [and Stephen Moss.](#)

## 498 **Funding**

499 This work was supported by a grant from the David and Lucile Packard Foundation to RPK (grant  
500 2014-39827). The funders had no role in study design, data collection and analysis, decision to  
501 publish, or preparation of the manuscript.

## 502 **Author Contributions**

503 Conceived and designed the experiments: JL O'Donnell, RP Kelly, AO Shelton; Collected the data:  
504 JL O'Donnell, NC Lowell, GD Williams, RP Kelly, AO Shelton, JF Samhouri; Conducted the  
505 analyses: JL O'Donnell; Wrote the first draft: JL O'Donnell; Edited the manuscript: JL O'Donnell,  
506 AO Shelton, RP Kelly, JF Samhouri, GD Williams, NC Lowell

## 507 **Ethics Statement**

508 The authors declare no conflict of interest. Consistent with the public trust doctrine, waters of the  
509 US are public, and therefore no permit was necessary to conduct this research (see Illinois Central  
510 Railroad v. Illinois, 146 U.S. 387 (1892)).

## 511 **Data Availability**

### 512 **0.1 ~~Sequence Data~~**

#### 513 Sequence Data

514 All sequence ~~files and metadata~~ data are available from ~~EMBL~~ NCBI under BioProject PRJNA338801.  
515 Scripts to process raw sequence data into the contingency tables used for ecological analyses can be  
516 found at:  
517 <https://github.com/jimmyodonnell/banzai>

518

## 519 **0.1 ~~Project Repository~~**

### 520 Project Repository

521 The following components are available from the project repository on GitHub:

522 [https://github.com/jimmyodonnell/Carkeek\\_eDNA\\_grid](https://github.com/jimmyodonnell/Carkeek_eDNA_grid)

523 <https://doi.org/10.5281/zenodo.242976>

## 524 **0.0.1 ~~Sequencing Metadata~~**

### 525 Sequencing Metadata

526 Sequencing metadata is available in: `Data/metadata_spatial.csv`

## 527 **0.0.1 ~~Life History Data~~**

### 528 Life History Data

529 Life history data is available in: `Data/life_history.csv`

## 530 **0.0.1 ~~Analysis Scripts~~**

### 531 Analysis Scripts

532 All analyses were performed using scripts available in the Analysis subdirectory of the project's  
533 repository on GitHub.

## 534 **References**

535 Anderson, M. J., Crist, T. O., Chase, J. M., Vellend, M., Inouye, B. D., Freestone, A. L., Sanders,  
536 N. J., Cornell, H. V., Comita, L. S., Davies, K. F., Harrison, S. P., Kraft, N. J. B., Stegen, J. C.,  
537 and Swenson, N. G. (2011). Navigating the multiple meanings of beta diversity: A roadmap for  
538 the practicing ecologist. *Ecology Letters*, 14(1):19–28.

539 Bahram, M., Kõljalg, U., Courty, P. E., Diédhiou, A. G., Kjølner, R., Põlme, S., Ryberg, M., Veldre,  
540 V., and Tedersoo, L. (2013). The distance decay of similarity in communities of ectomycorrhizal  
541 fungi in different ecosystems and scales. *Journal of Ecology*, 101(5):1335–1344.

542 Barnes, M. A., Turner, C. R., Jerde, C. L., Renshaw, M. A., Chadderton, W. L., and Lodge, D. M.  
 543 (2014). Environmental conditions influence eDNA persistence in aquatic systems. *Environmental*  
 544 *Science and Technology*, 48(3):1819–1827.

545 Bell, T. (2010). Experimental tests of the bacterial distance–decay relationship. *The ISME Journal*,  
 546 4(11):1357–1365.

547 Burns, R. E. (1985). *The shape and form of Puget Sound*. Washington Sea Grant, Seattle, 1 edition.

548 Calvo, M. and Templado, J. (2004). Reproduction and development in a vermetid gastropod,  
 549 *Vermetus triquetrus*. *Invertebrate Biology*, 123(4):289–303.

550 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L.  
 551 (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10:421.

552 Chamberlain, S. a. and Szöcs, E. (2013). taxize: taxonomic search and retrieval in R. *F1000Research*,  
 553 2(0):191.

554 Chamberlain, S. A., Szöcs, E., Boettiger, C., Ram, K., Bartomeus, I., Foster, Z., and O’Donnell,  
 555 J. L. (2016). taxize: Taxonomic information from around the web. R package.

556 Chust, G., Chave, J., Condit, R., Aguilar, S., Lao, S., and Perez, R. (2006). Determinants and  
 557 spatial modeling of tree beta-diversity in a tropical forest landscape in Panama. *Journal of*  
 558 *Vegetation Science*, 17(1):83–92.

559 Chust, G., Irigoien, X., Chave, J., and Harris, R. P. (2013). Latitudinal phytoplankton distribution  
 560 and the neutral theory of biodiversity. *Global Ecology and Biogeography*, 22(5):531–543.

561 Coissac, E. (2012). OligoTag: A Program for Designing Sets of Tags for Next-Generation Sequencing  
 562 of Multiplexed Samples. In Pompanon, F. and Bonin, A., editors, *Data Production and Analysis in*  
 563 *Population Genomics SE - 2*, volume 888 of *Methods in Molecular Biology*, pages 13–31. Humana  
 564 Press.

565 Condit, R. (2002). Beta-Diversity in Tropical Forest Trees. *Science*, 295(5555):666–669.

566 Cowart, D. a., Pinheiro, M., Mouchel, O., Maguer, M., Grall, J., Miné, J., and Arnaud-Haond, S.  
567 (2015). Metabarcoding Is Powerful yet Still Blind: A Comparative Analysis of Morphological and  
568 Molecular Surveys of Seagrass Communities. *Plos One*, 10(2):e0117562.

569 de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le  
570 Bescot, N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J.-M., Bittner,  
571 L., Chaffron, S., Dunthorn, M., Engelen, S., Flegontova, O., Guidi, L., Horák, A., Jaillon, O.,  
572 Lima-Mendez, G., Lukeš, J., Malviya, S., Morard, R., Mulot, M., Scalco, E., Siano, R., Vincent,  
573 F., Zingone, A., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Acinas, S. G., Bork, P.,  
574 Bowler, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S.,  
575 Raes, J., Sieracki, M. E., Speich, S., Stemmann, L., Sunagawa, S., Weissenbach, J., Wincker, P.,  
576 and Karsenti, E. (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science*, 348(6237).

577 Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., Taberlet, P., Deagle, B. E., Jarman, S. N.,  
578 and Coissac, E. (2014). DNA metabarcoding and the cytochrome c oxidase subunit I marker :  
579 not a perfect match DNA metabarcoding and the cytochrome c oxidase subunit I marker : not a  
580 perfect match. (September).

581 Deiner, K. and Altermatt, F. (2014). Transport distance of invertebrate environmental DNA in a  
582 natural river. *PLoS ONE*, 9(2).

583 Dethier, M. N. (2010). Overview of the ecology of Puget Sound beaches. In Shipman, H., Dethier,  
584 M. N., Gelfenbaum, G., Fresh, K. L., and Dinicola, R. S., editors, *Puget Sound Shorelines and*  
585 *the Impacts of Armoring—Proceedings of a State of the Science Workshop*, page 262.

586 Drummond, A. J., Newcomb, R. D., Buckley, T. R., Xie, D., Dopheide, A., Potter, B. C., Heled, J.,  
587 Ross, H. A., Tooman, L., Grosser, S., Park, D., Demetras, N. J., Stevens, M. I., Russell, J. C.,  
588 Anderson, S. H., Carter, A., and Nelson, N. (2015). Evaluating a multigene environmental DNA  
589 approach for biodiversity assessment. *GigaScience*, 4(1):46.

590 Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*,  
591 26(19):2460–2461.

592 Evans, N. T., Olds, B. P., Renshaw, M. A., Turner, C. R., Li, Y., Jerde, C. L., Mahon, A. R.,

593 Pfrender, M. E., Lamberti, G. A., and Lodge, D. M. (2016). Quantification of mesocosm fish and  
 594 amphibian species diversity via environmental DNA metabarcoding. *Molecular Ecology Resources*,  
 595 16(1):29–41.

596 Ficetola, G. F., Miaud, C., Pompanon, F., and Taberlet, P. (2008). Species detection using envi-  
 597 ronmental DNA from water samples. *Biology letters*, 4(4):423–425.

598 Fonseca, V. G., Carvalho, G. R., Nichols, B., Quince, C., Johnson, H. F., Neill, S. P., Lamshead,  
 599 J. D., Thomas, W. K., Power, D. M., and Creer, S. (2014). Metagenetic analysis of patterns of  
 600 distribution and diversity of marine meiobenthic eukaryotes. *Global Ecology and Biogeography*,  
 601 23(11):1293–1302.

602 Froese, R. and Pauly, D. (2016). FishBase.

603 Guardiola, M., Uriz, M. J., Taberlet, P., Coissac, E., Wangenstein, O. S., and Turon, X.  
 604 (2015). Deep-Sea, Deep-Sequencing: Metabarcoding Extracellular DNA from Sediments of Ma-  
 605 rine Canyons. *PLOS ONE*, 10(10):e0139633.

606 Guardiola, M., Wangenstein, O. S., Taberlet, P., Coissac, E., Uriz, M. J., and Turon, X. (2016).  
 607 Spatio-temporal monitoring of deep-sea communities using metabarcoding of sediment DNA and  
 608 RNA. *PeerJ*, 4:e2807.

609 Guo, H., Chamberlain, S. A., Elhaik, E., Jalli, I., Lynes, A. R., Marczak, L., Sabath, N., Vargas,  
 610 A., Więski, K., Zelig, E. M., and Pennings, S. C. (2015). Geographic variation in plant commu-  
 611 nity structure of salt marshes: Species, functional and phylogenetic perspectives. *PLoS ONE*,  
 612 10(5):e0127781.

613 Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., and Goodman, R. M. (1998). Molecular  
 614 biological access to the chemistry of unknown soil microbes: a new frontier for natural products.  
 615 *Chemistry & Biology*, 5(10):R245–R249.

616 Hijmans, R. J. (2016). geosphere: Spherical Trigonometry.

617 Hubbell, S. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography.*, volume 32.



618 Iverson, V., Morris, R. M., Frazar, C. D., Berthiaume, C. T., Morales, R. L., and Armbrust,  
 619 E. V. (2012). Untangling Genomes from Metagenomes: Revealing an Uncultured Class of Marine  
 620 Euryarchaeota. *Science*, 335(6068):587–590.

621 Kaufman, L. and Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster*  
 622 *Analysis*.

623 Kelly, R. P., Closek, C. J., O'Donnell, J. L., Kralj, J. E., Shelton, A. O., and Samhour, J. F.  
 624 (2016a). Genetic and manual survey methods yield different and complementary views of an  
 625 ecosystem. *Frontiers in Marine Science*, 3:283.

626 Kelly, R. P., O'Donnell, J. L., Lowell, N. C., Shelton, A. O., Samhour, J. F., Hennessey, S. M.,  
 627 Feist, B. E., and Williams, G. D. (2016b). Genetic signatures of ecological diversity along an  
 628 urbanization gradient. *PeerJ*, 4:e2444.

629 Klymus, K. E., Richter, C. A., Chapman, D. C., and Paukert, C. (2015). Quantification of eDNA  
 630 shedding rates from invasive bighead carp *Hypophthalmichthys nobilis* and silver carp *Hypoph-*  
 631 *thalmichthys molitrix*. *Biological Conservation*, 183:77–84.

632 Kozloff, E. N. (1973). *Seashore life of Puget Sound, the Strait of Georgia, and the San Juan*  
 633 *Archipelago*. University of Washington Press, Seattle.

634 Lallias, D., Hiddink, J. G., Fonseca, V. G., Gaspar, J. M., Sung, W., Neill, S. P., Barnes, N.,  
 635 Ferrero, T., Hall, N., Lambshead, P. J. D., Packer, M., Thomas, W. K., and Creer, S. (2015).  
 636 Environmental metabarcoding reveals heterogeneous drivers of microbial eukaryote diversity in  
 637 contrasting estuarine ecosystems. *The ISME journal*, 9(5):1208–21.

638 Levin, S. A. (1992). The problem of pattern and scale in ecology. *Ecology*, 73(6):1943–1967.

639 Longmire, J. L., Maltbie, M., and Baker, R. J. (1997). Use of lysis buffer in DNA isolation and its  
 640 implication for museum collections. *Museum of Texas Tech University*, 163.

641 Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2016). *cluster: Cluster*  
 642 *Analysis Basics and Extensions*.

643 Magurran, A. E. (2003). *Measuring Biological Diversity*. Wiley.

644 Mahé, F., Rognes, T., Quince, C., de Vargas, C., and Dunthorn, M. (2014). Swarm: robust and  
 645 fast clustering method for amplicon-based studies. *PeerJ*, 2:e593.

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

648 Martiny, J. B. H., Eisen, J. A., Penn, K., Allison, S. D., and Horner-Devine, M. C. (2011). Drivers of  
 649 bacterial beta diversity depend on spatial scale. *Proceedings of the National Academy of Sciences*,  
 650 108(19):7850–7854.

651 Nekola, J. C. and White, P. S. (1999). The distance decay of similarity in biogeography and ecology.  
 652 *Journal of Biogeography*, 26(4):867–878.

653 O’Donnell, J. L., Kelly, R. P., Lowell, N. C., and Port, J. A. (2016). Indexed PCR Primers Induce  
 654 Template-Specific Bias in Large-Scale DNA Sequencing Studies. *PLOS ONE*, 11(3):e0148698.

655 Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R.,  
 656 O’Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., and Wagner, H.  
 657 (2016). *vegan: Community Ecology Package*.

658 Parr, C. S., Wilson, N., Leary, P., Schulz, K. S., Lans, K., Walley, L., Hammock, J. A., Goddard,  
 659 A., Rice, J., Studer, M., Holmes, J. T. G., and Corrigan, R. J. (2014). The Encyclopedia of  
 660 Life v2: Providing Global Access to Knowledge About Life on Earth. *Biodiversity Data Journal*,  
 661 2(2):e1079.

662 Phillips, N. E. and Shima, J. S. (2010). Reproduction of the vermetid gastropod *dendropoma*  
 663 *maximum* (Sowerby, 1825) in Moorea, French Polynesia. *Journal of Molluscan Studies*, 76(2):133–  
 664 137.

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

668 R Core Team (2016). *R: A Language and Environment for Statistical Computing*.

669 Ray, G. C. (1988). Ecological diversity in coastal zones and oceans. In Wilson, E. O. and Peter,  
 670 F. M., editors, *Biodiversity*, chapter 4. National Academies Press (US), Washington, DC.

671 Renshaw, M. A., Olds, B. P., Jerde, C. L., McVeigh, M. M., and Lodge, D. M. (2015). The  
 672 room temperature preservation of filtered environmental DNA samples and assimilation into a  
 673 phenol–chloroform–isoamyl alcohol DNA extraction. *Molecular Ecology Resources*, 15(1):168–176.

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

676 Sassoubre, L. M., Yamahara, K. M., Gardner, L. D., Block, B. A., and Boehm, A. B. (2016).  
 677 Quantification of Environmental DNA (eDNA) Shedding and Decay Rates for Three Marine  
 678 Fish. *Environmental Science & Technology*, 50(19):10456–10464.

679 Schnell, I. B., Bohmann, K., and Gilbert, M. T. P. (2015). Tag jumps illuminated - reducing  
 680 sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*,  
 681 pages n/a–n/a.

682 Shelton, A. O., O'Donnell, J. L., Samhour, J. F., Lowell, N., Williams, G. D., and Kelly, R. P.  
 683 (2016). A framework for inferring biological communities from environmental DNA. *Ecological*  
 684 *Applications*, 26(6):1645–1659.

685 Shogren, A. J., Tank, J. L., Andruszkiewicz, E. A., Olds, B., Jerde, C., and Bolster, D. (2016).  
 686 Modelling the transport of environmental DNA through a porous substrate using continuous  
 687 flow-through column experiments. *Journal of The Royal Society Interface*, 13(119):423–425.

688 Simpson, E. H. (1949). Measurement of diversity. *Nature*, 163(688).

689 Spiess, A.-N. (2014). propagate: Propagation of Uncertainty.

690 Strathmann, M. F. (1987). *Reproduction and Development of Marine Invertebrates of the Northern*  
 691 *Pacific Coast: Data and Methods for the Study of Eggs, Embryos, and Larvae*. University of  
 692 Washington Press, Seattle.

693 Strathmann, M. F. and Strathmann, R. R. (2006). A Vermetid Gastropod with Complex Intracap-  
694 sular Cannibalism of Nurse Eggs and Sibling Larvae and a High Potential for Invasion. *Pacific*  
695 *Science*, 60(1):97–108.

696 Strickland, R. M. (1983). *The Fertile Fjord: Plankton in Puget Sound*. University of Washington  
697 Press, Seattle.

698 Strickler, K. M., Fremier, A. K., and Goldberg, C. S. (2015). Quantifying effects of UV-B, temper-  
699 ature, and pH on eDNA degradation in aquatic microcosms. *Biological Conservation*, 183:85–92.

700 Thomsen, P. F., Kielgast, J., Iversen, L. L., Møller, P. R., Rasmussen, M., and Willerslev, E. (2012).  
701 Detection of a Diverse Marine Fish Fauna Using Environmental DNA from Seawater Samples.  
702 *PLoS ONE*, 7(8):1–9.

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

705 Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M.,  
706 Solovyev, V. V., Rubin, E. M., Rokhsar, D. S., and Banfield, J. F. (2004). Community structure  
707 and metabolism through reconstruction of microbial genomes from the environment. *Nature*,  
708 428(6978):37–43.

709 Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. a., Wu, D.,  
710 Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W.,  
711 Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch,  
712 C., Rogers, Y.-H., and Smith, H. O. (2004). Environmental genome shotgun sequencing of the  
713 Sargasso Sea. *Science*, 304(5667):66–74.

714 Wetzel, C. E., de Bicudo, D. C., Ector, L., Lobo, E. A., Soininen, J., Landeiro, V. L., and Bini,  
715 L. M. (2012). Distance Decay of Similarity in Neotropical Diatom Communities. *PLoS ONE*,  
716 7(9):e45071.

717 Yang, Z. and Khangaonkar, T. (2010). Multi-scale modeling of Puget Sound using an unstructured-  
718 grid coastal ocean model: From tide flats to estuaries and coastal waters. *Ocean Dynamics*,  
719 60(6):1621–1637.

720 Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: A fast and accurate Illumina  
721 Paired-End reAd mergeR. *Bioinformatics*, 30(5):614–620.

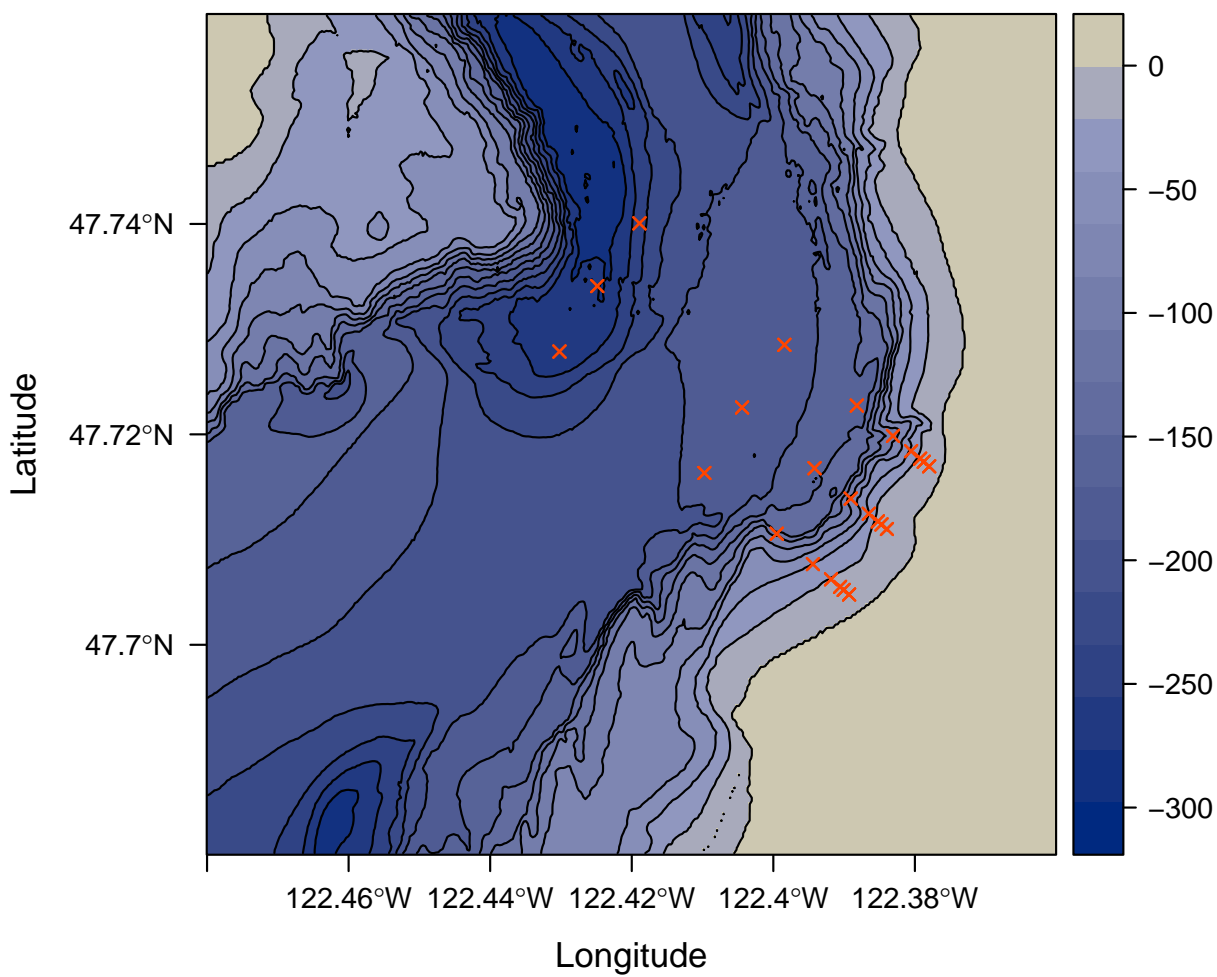


Figure 1: Map of study area. Depth in meters below sea level is indicated by shading and 25 meter contours. Sampled locations are indicated by red points.

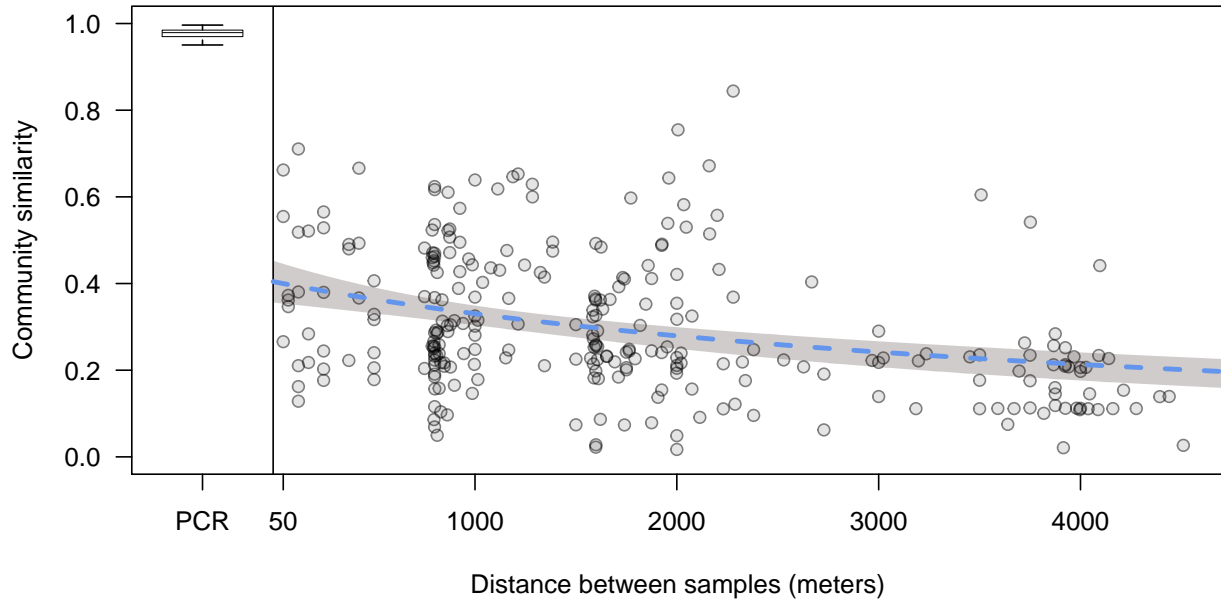


Figure 2: Distance decay relationship of environmental DNA communities. Each point represents the Bray-Curtis similarity of a site sampled along three parallel transects comprising a 3000 by 4000 meter grid. Blue dashed line represents fit of a nonlinear least squares regression (see Methods), and shading denotes the 95% confidence interval. Boxplot is comparisons within-sample across PCR replicates, separated by a vertical line at zero, where the central line is the median, the box encompasses the interquartile range, and the lines extend to 1.5 times the interquartile range. Boxplot outliers are omitted for clarity.



Figure 3: Cluster membership of sampled sites. Distance from onshore starting point is log scaled. Sites are colored and labeled by their assignment to a cluster by PAM analysis for number of clusters (K) chosen based on a priori expectations (2) and mean silhouette width (8).



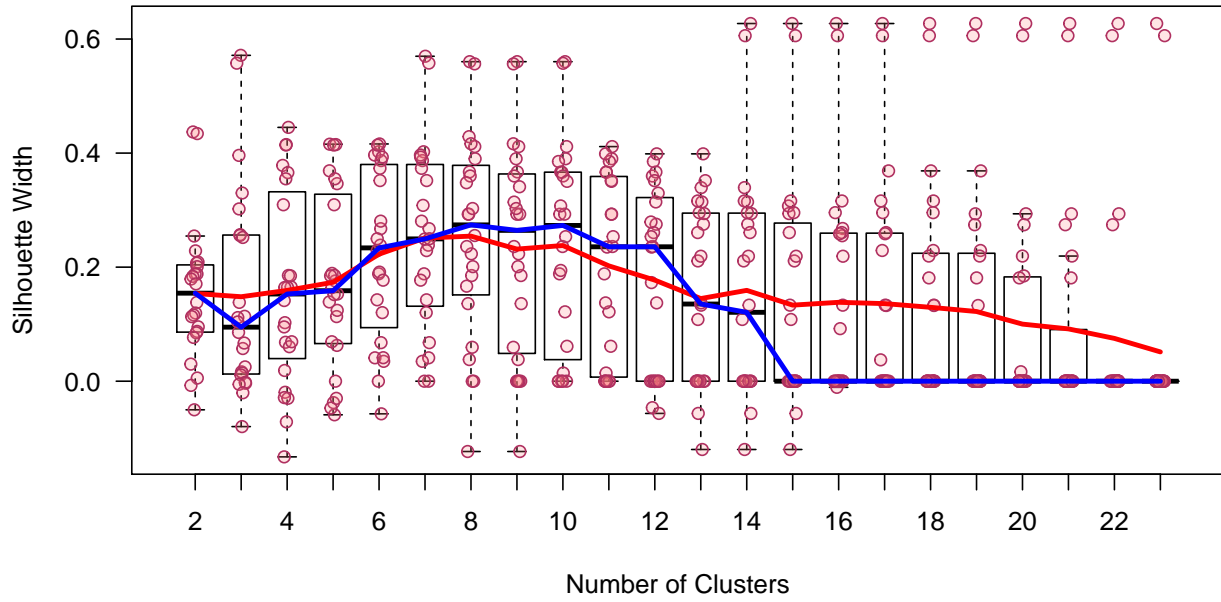


Figure 4: Silhouette widths from PAM analysis. Points are the width of the PAM silhouette of each sample at each number of clusters ( $K$ ). Red line is the mean, blue line is the median. Boxes encompass the interquartile range with a line at the median, and the whiskers extend to 1.5 times the interquartile range. Boxplot outliers are omitted for clarity.

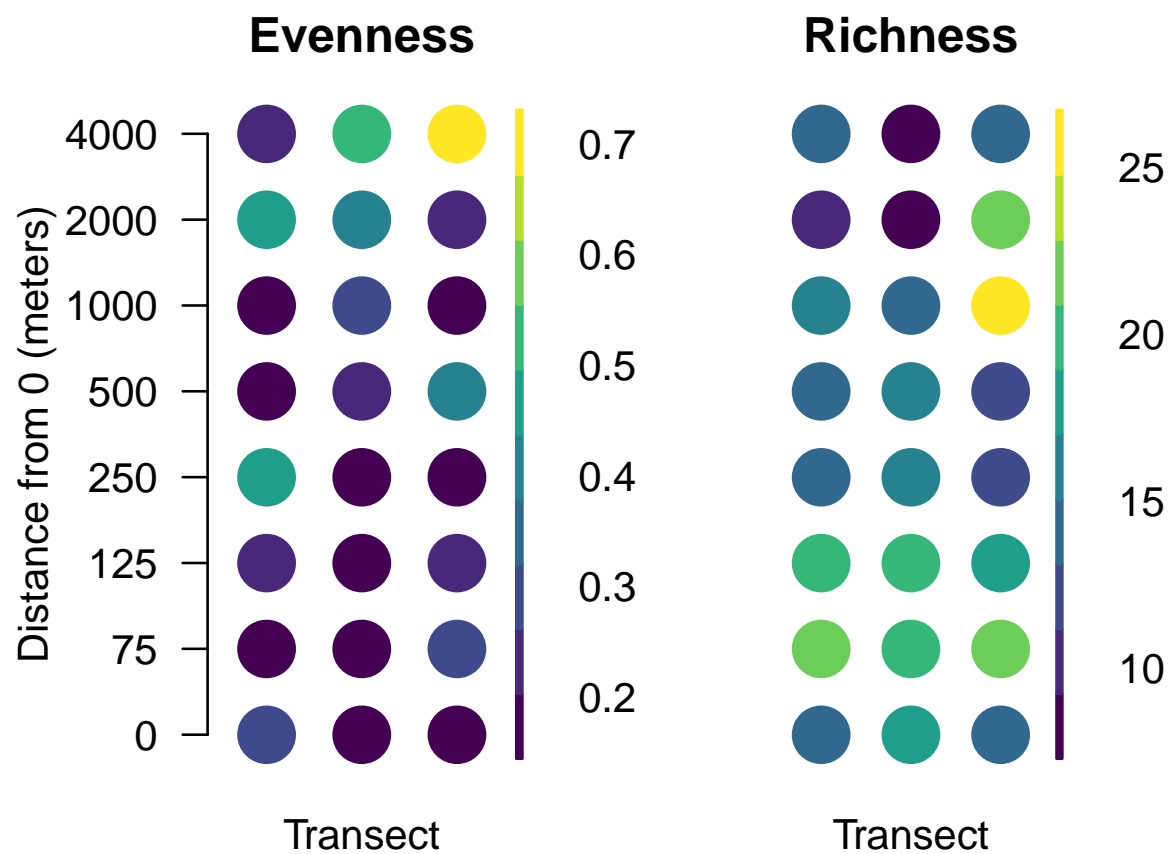


Figure 5: Aggregate measures of diversity at each sample site. Data are rarefied counts of mitochondrial 16S sequences collected from 3 parallel transects in Puget Sound, Washington, USA. Evenness (left) is the probability that two sequences drawn at random are different; richness (right) represents the total number of unique sequences from that location.

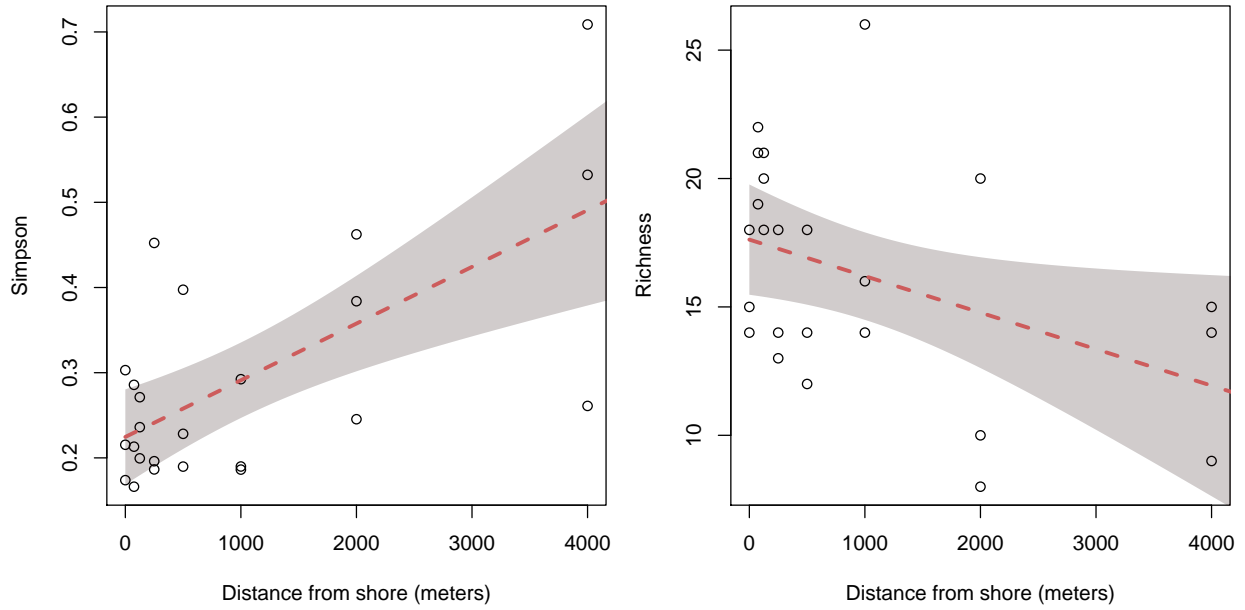


Figure 6: Aggregate diversity metrics of each site plotted against distance from shore. Both Simpson's Index (left) and richness (right) are shown, and have been computed from the mean abundance of unique DNA sequences found across 4 PCR replicates at each of 24 sites. Lines and bands illustrate the fit and 95

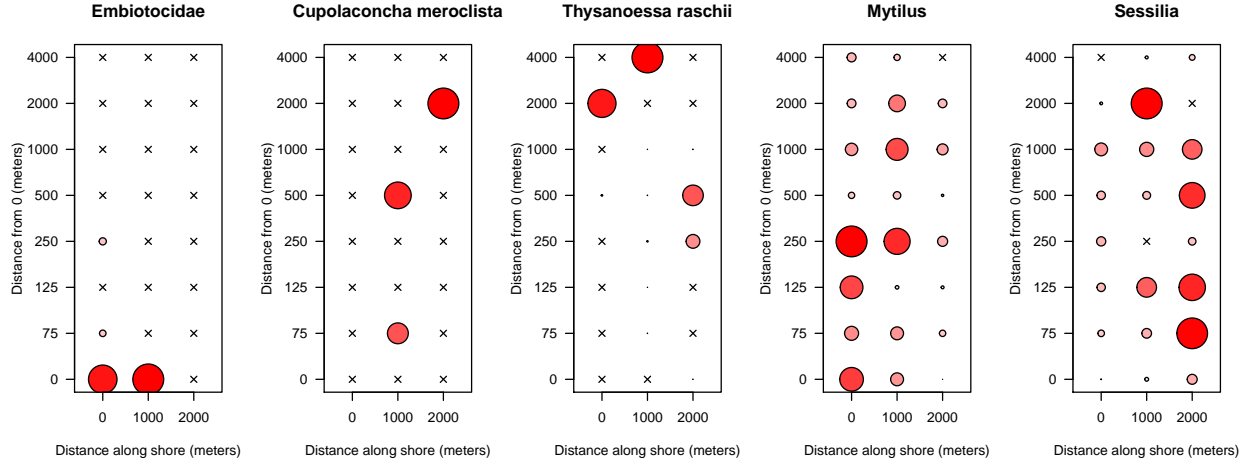


Figure 7: Distribution of eDNA from select taxa. Circles are colored and scaled by the proportion of that taxon's maximum proportional abundance. That is, the largest circle is the same size in each of the panels, and occurs where that taxon contributed the greatest proportional abundance of reads to that sample.

## 723 Supplemental Material

### 724 Methods

#### 725 Bioinformatics

726 Reads passing the preliminary Illumina quality filter were demultiplexed on the basis of the adapter  
727 index sequence by the sequencing facility. We used fastqc to assess the fastq files output from the  
728 sequencer for low-quality indications of a problematic run. Forward and reverse reads were merged  
729 using PEAR v0.9.6 Zhang et al. (2014) and discarded if more than 0.01 of the bases were uncalled.  
730 If a read contained two consecutive base calls with quality scores less than 15 (i.e. probability of  
731 incorrect base call = 0.0316), these bases and all subsequent bases were removed from the read.  
732 Paired reads for which the probability of matching by chance alone exceeded 0.01 were not assembled  
733 and omitted from the analysis. Assembled reads were discarded if assembled sequences were not  
734 between 50 and 168 bp long, or if reads did not overlap by at least 100 bp.

735 We used vsearch v2.1.1 (Rognes et al., 2016) to discard any merged reads for which the sum of the  
736 per-base error probabilities was greater than 0.5 (“expected errors”) Edgar (2010). Sequences were  
737 demultiplexed on the basis of the primer index sequence at base positions 4-9 at both ends using the  
738 programming language AWK. Primer sequences were removed using cutadapt v1.7.1 Martin (2011),  
739 allowing for 2 mismatches in the primer sequence. Identical duplicate sequences were identified,  
740 counted, and removed in python to speed up subsequent steps by eliminating redundancy, and  
741 sequences occurring only once were removed. We checked for and removed any sequence likely to be  
742 a PCR artifact due to incomplete extension and subsequent mis-priming using a method described  
743 by Edgar (2010) and implemented in vsearch v2.0.2. Sequences were clustered into operational  
744 taxonomic units (OTUs) using the single-linkage clustering method implemented by swarm version  
745 2.1.1 with a local clustering threshold (d) of 1 and fastidious processing (Mahé et al., 2014).

746 Cross-contamination of environmental, DNA, or PCR samples can result in erroneous inference  
747 about the presence of a given DNA sequence in a sample. However, other processes can contribute  
748 to the same signature of contamination. For example, errors during oligonucleotide synthesis or  
749 sequencing of the indexes could cause reads to be erroneously assigned to samples. The frequency  
750 of such errors can be estimated by counting the occurrence of sequences known to be absent from

751 a given sample, and of reads that do not contain primer index sequences in the expected position  
752 or combinations. These occurrences indicate an error in the preparation or sequencing procedures.  
753 We estimated a rate of incorrect sample assignment by calculating the maximum rate of occur-  
754 rence of index sequences combinations we did not actually use, as well as the rates of cross-library  
755 contamination by counting occurrences of primer sequences from 12S amplicons prepared in a lab  
756 more than 1000 kilometers away, but pooled and sequenced alongside our samples. This represents  
757 a general minimum rate at which we can expect that sequences from one environmental sample  
758 could be erroneously assigned to another, and so we considered for further analysis only those reads  
759 occurring with greater frequency than this across the entire dataset.

760 We checked for experimental error by evaluating the Bray-Curtis similarity (1 - Bray-Curtis  
761 dissimilarity) among replicate PCRs from the same DNA sample. We calculated the mean and  
762 standard deviation across the dataset, and excluded any PCR replicates for which the similarity  
763 between itself and the other replicates was less than 1.5 standard deviations from the mean.

764 To account for variation in the number of sequencing reads (sequencing depth) recovered per  
765 sample, we rarefied the within-sample abundance of each OTU by the minimum sequencing depth  
766 (Oksanen et al., 2016).

767 Because each step in this workflow is sensitive to contamination, it is possible that some se-  
768 quences are not truly derived from the environmental sample, and instead represent contamination  
769 during field sampling, filtration, DNA extraction, PCR, fragment size selection, quantitation, se-  
770 quencing adapter ligation, or the sequencing process itself. We take the view that contaminants  
771 are unlikely to manifest as sequences in the final dataset in consistent abundance across replicates;  
772 indeed, our data show that the process from PCR onward is remarkably consistent. Thus, after  
773 scaling to correct for sequencing depth variation, we calculated from our data the maximum number  
774 of sequence counts for which there is turnover in presence-absence among PCR replicates within an  
775 environmental sample. We use this number to determine a conservative minimum threshold above  
776 which we can be confident that counts are consistent among replicates and not of spurious origin,  
777 and exclude from further analysis observations where the mean abundance across PCR replicates  
778 within samples does not reach this threshold. For further analyses we use the mean abundance  
779 across PCR replicates for each of the 24 environmental samples.

780 In order to determine the most likely taxon from which each sequence originated, the representa-

781 tive sequence from each OTU was then queried against the NCBI nucleotide collection (GenBank;  
782 version October 7, 2015; 32,827,936 sequences) using the blastn command line utility (Camacho  
783 et al., 2009). In order to maximize the accuracy of this computationally intensive step, we imple-  
784 mented a nested approach whereby each sequence was first queried using strict parameters (e-value  
785 = 5e-52), and if no match was found, the query was repeated with decreasingly strict e-values (5e-48  
786 5e-44 5e-40 5e-36 5e-33 5e-29 5e-25 5e-21 5e-17 5e-13). Other parameters were unchanged among  
787 repetitions (word size: 7; maximum matches: 1000; culling limit: 100; minimum percent identity:  
788 0). Each query sequence can be an equally good match to multiple taxa either because of invariabil-  
789 ity among taxa or errors in the database (e.g. human sequences are commonly attributed to other  
790 organisms when they in fact represent lab contamination). In order to guard against these spurious  
791 results, we used an algorithm to find the lowest common taxon for at least 80% of the matched  
792 taxa, implemented in the R package taxize 0.7.8 (Chamberlain and Szöcs, 2013; Chamberlain et al.,  
793 2016). Similarly, we repeated analyses using the dataset consolidated at the same taxonomic rank  
794 across all queries, for the rank of both family and order.

## 795 **Alternative distance decay model formulations**

796 **Linear:** We fit a straight line through the points after log-transforming the spatial distances  
797 to estimate the intercept and slope. This model ignores the bounds of our response variable of  
798 community similarity.

799 **Michaelis-Menten:** We fit a Michaelis-Menten-like curve to our data. Our formulation can be  
800 thought of as a modification of the Michaelis-Menten equation, but with the addition of a parameter  
801 in the numerator which modifies the intercept.

$$y = \frac{AB + Cx}{B + x} \quad (2)$$

802 Where  $C$  is the asymptote of minimum similarity. This formulation allows us to estimate the  
803 maximum similarity in the system, and the rate at which it is achieved. If the value of the parameter  
804 ( $AB$ ) is 0 (i.e. if the intercept is 0), the form is identical to the Michaelis-Menten equation:

$$y = \frac{Cx}{B+x} \quad (3)$$

805 This is conceptually satisfying in that a fit through [0,1] reflects the theoretical expectation that  
806 samples at zero distance from one another are necessarily identical. Given an efficient sampling  
807 technique, replicate samples taken at the same position in space should be identical, and thus the  
808 intercept of the regression of similarity against distance should be 1, and deviation from 1 is an  
809 indicator of the efficiency of the sampling method.

810 Finally, we considered a model which estimates an asymptote as the total change in similarity  
811 ( $D$ ):

$$y = \frac{A + Dx}{B + x} \quad (4)$$

812 However, this model failed to converge and produced uninformative estimates of all parameters.



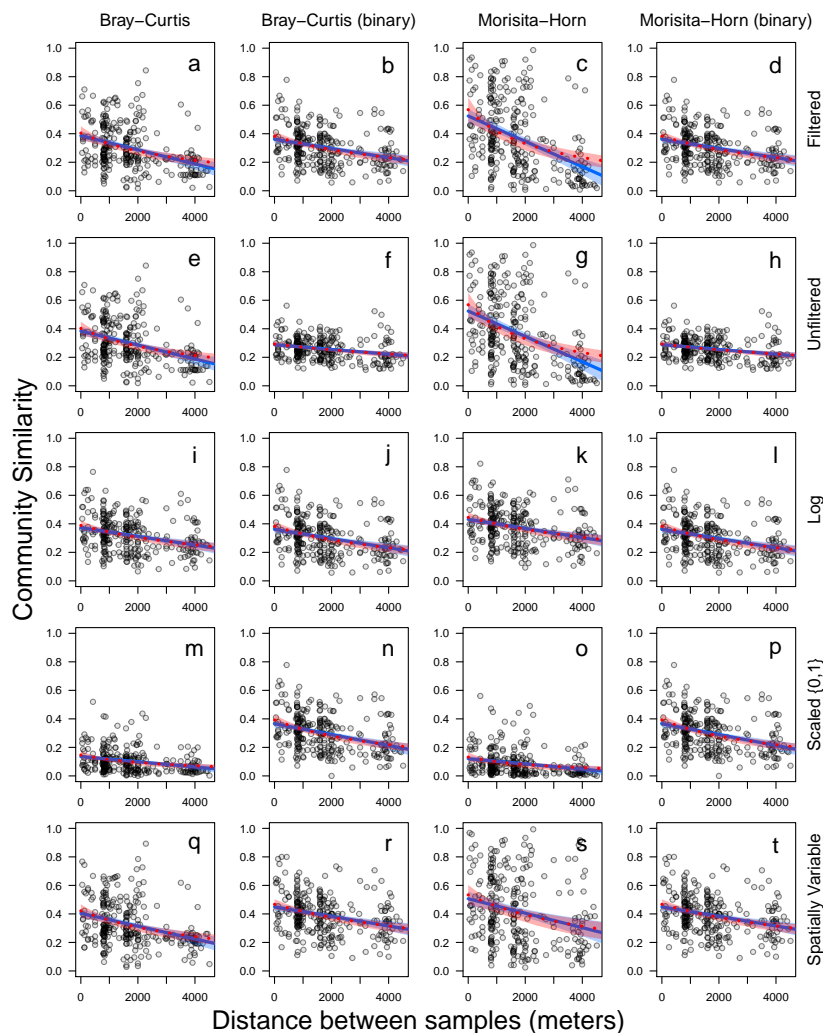


Figure 8: Distance decay relationship of environmental DNA communities using a variety of models, metrics, and data subsets. Each point represents the similarity of a site sampled along three parallel transects comprising a 3000 by 4000 meter grid. Each row of plots represents a different data subset indicated in the right margin, including the final filtered data reported in the main text (a-d), the unfiltered data including all rare OTUs (e-h), log-transformed ( $\log(x+1)$ ) data (i-l), OTU abundance scaled relative to within-taxon maximum (m-p), and exclusion of OTUs found at only one site (q-t). Columns indicate the similarity index used (Bray Curtis or Morisita-Horn) and whether the input was full abundance data or binary (0,1) transformed data. Lines and bands illustrate the fit and 95% confidence interval of both the main nonlinear model (red, dashed line) and a simple linear model (blue, solid line). Results using the Jaccard distance are omitted because of its similarity to Bray-Curtis.

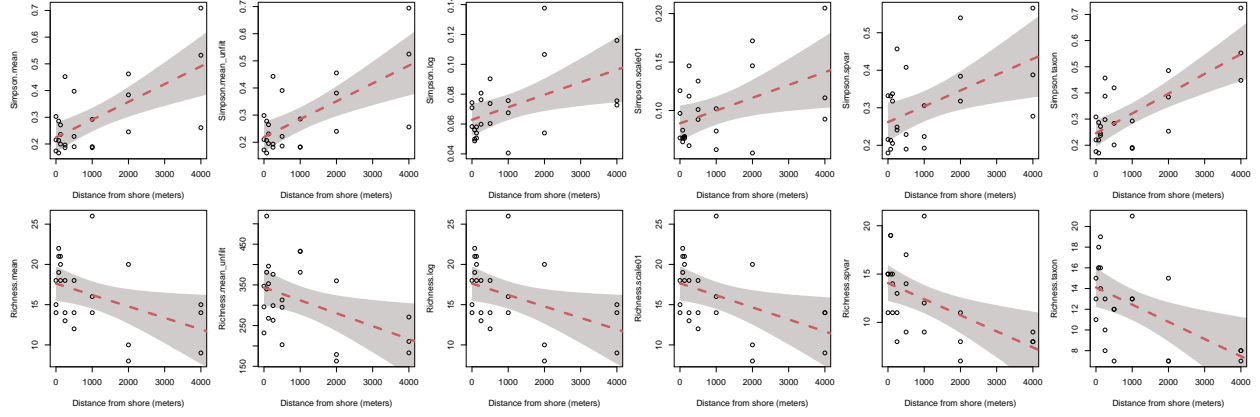


Figure 9: Aggregate diversity metrics of each site plotted against distance from shore. Both Simpson's Index (top) and richness (bottom) are shown for a variety of data subsets and transformations (left to right: mean, unfiltered mean,  $\log(x + 1)$ , transformed, scaled, spatially variable, and taxon clustered). Lines and bands illustrate the fit and 95% confidence interval of a linear model. See methods text for detailed data descriptions.

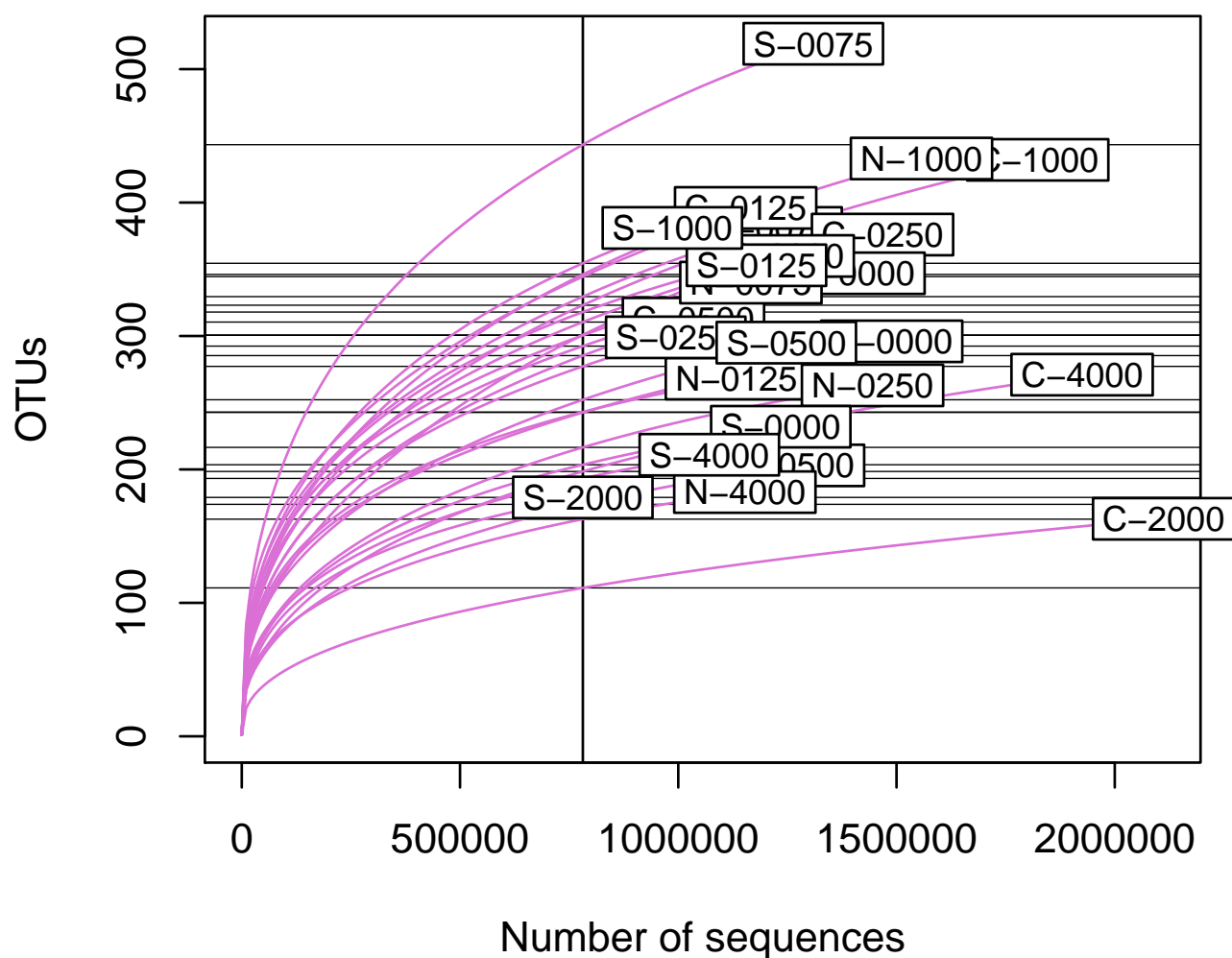


Figure 10: Accumulation of OTUs from 24 environmental samples using randomized rarefaction. Four replicate PCRs were conducted using DNA each environmental sample and independently sequenced, but these are collapsed here to illustrate a single representation of richness. Sample names indicate the position in the sampling grid: south (S), central (C), or north (N), followed by the distance along the transect, in meters (0, 75, 125, 250, 500, 1000, 2000, 4000). Vertical line indicates the minimum combined number of sequence reads per sample. Horizontal lines indicate OTU richness for each sample at the minimum combined number of sequence reads.