

Spatial distribution of environmental DNA in a nearshore marine habitat

James L. O'Donnell^{*1}, Ryan P. Kelly¹, Andrew O. Shelton², Jameal F. Samhouri³, Natalie C. Lowell^{1,4}, and Gregory D. Williams⁵

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

²Earth 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

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

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

⁵Pacific 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 11, 2017

Keywords

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

^{*}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 PCR primers that target a diverse assemblage of metazoans, we amplified a region of mitochondrial 16S 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 to be more similar than those that are farther apart (Nekola and White, 1999). This decrease in community similarity with increasing spatial separation is called distance decay and has been

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

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

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

48 Macroorganisms shed DNA-containing cells into the environment (environmental DNA or eDNA)
49 that can be sampled in the same way (Ficetola et al., 2008; Thomsen et al., 2012). Potentially, eDNA
50 methods allow a broad swath of macroorganisms to be surveyed from basic environmental samples.
51 However, the accuracy and reliability of indirect estimates of macroorganismal abundance has been
52 debated because the entire organisms are not contained within the sample (Coward et al., 2015).
53 Concern surrounding eDNA methods is rooted in uncertainty about the attributes of eDNA in the
54 environment relative to actual organisms (Shelton et al., 2016; Evans et al., 2016). Basic questions
55 such as how long DNA can persist in that environment and how far DNA can travel remain largely
56 unknown (but see Klymus et al. (2015); Turner et al. (2015); Strickler et al. (2015); Deiner and
57 Altermatt (2014)) and impede inference about local organismal presence from an environmental
58 sample. As a result, estimating the spatial and temporal resolution of eDNA studies in the field is

a key step in making these methods practical.

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

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

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

Puget Sound is a deep, narrow fjord in Washington, USA, where a narrow band of shallow

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

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

diversity measures of eDNA communities that have been widely used to reveal broad scale patterns based on macrobiota in many ecological systems. Finally, we identify the taxa represented in the eDNA communities, which span a range of life-history characteristics, and we expect that the spatial distribution of eDNA will most closely resemble the distribution of adults in taxa with low dispersal potential.

Methods

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

Environmental Sampling

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

Filtration

One liter from each water sample was filtered in the lab on a clean polysulfone vacuum filter holder fitted with a 47 millimeter diameter cellulose acetate membrane with 0.45 micrometer pores.

145 Filter membranes were moved into 900 microliters of Longmire buffer (Longmire et al., 1997) using
146 clean forceps and stored at room temperature (Renshaw et al., 2015). To test for the extent of
147 contamination attributable to laboratory procedures, we filtered three replicate 1 liter samples of
148 deionized water. These samples were treated identically to the environmental samples throughout
149 the remaining protocols.

150 DNA Purification

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

163 PCR Amplification

164 We chose a primer set that amplifies an approximately 115 base pair (bp) region of the mitochon-
165 drial 16S rRNA gene in at least 10 metazoan phyla from this habitat, excludes non-metazoans, and
166 resolves taxonomy to the family level in most cases using a public sequence database (Kelly et al.,
167 2016a). We used a two-step polymerase chain reaction (PCR) protocol described by O'Donnell
168 et al. (2016) to generate 4 replicate products from each DNA sample. In the first set of reactions,
169 primers were identical in every reaction (forward: AGTTACYYTAGGGATAACAGCG; reverse:
170 CCGGTCTGAACTCAGATCAYGT); primers in the second set of reactions included these same
171 sequences but with 3 variable nucleotides (NNN) and an index sequence on the 5' end (see Sequenc-
172 ing Metadata). We used the program OligoTag (Coissac, 2012) to generate 30 unique 6-nucleotide

index sequences differing by a minimum Hamming distance of 3 (see Sequencing Metadata). Indexed primers were assigned to samples randomly, with the identical index sequence on the forward and reverse primer to avoid errors associated with dual-indexed multiplexing (Schnell et al., 2015). In a UV-sterilized hood, we prepared 25 microliter reactions containing 18.375 microliters ultra-pure water, 2.5 microliters 10x buffer, 0.625 microliters deoxynucleotide solution (8 millimolar), 1 microliter each forward and reverse primer (10 micromolar, obtained lyophilized from Integrated DNA Technologies (Coralville, IA, USA)), 0.25 microliters Qiagen HotStar Taq polymerase, and 1.25 microliter genomic eDNA template at 1:100 dilution in ultrapure water. PCR thermal profiles began with an initialization step (95 °C; 15 min) followed by cycles (40 and 20 for the first and second reaction, respectively) of denaturation (95 °C; 15 sec), annealing (61 °C; 30 sec), and extension (72 °C; 30 sec). 20 identical PCRs were conducted from each DNA extract using non-indexed primers; these were pooled into 4 groups of 5 in order to ensure ample template for the subsequent PCR with indexed primers. In order to isolate the fragment of interest from primer dimer and other spurious fragments generated in the first PCR, we used the AxyPrep Mag FragmentSelect-I kit with solid-phase reversible immobilization (SPRI) paramagnetic beads at 2.5x the volume of PCR product (Axygen BioSciences, Corning, NY, USA). A 1:5 dilution in ultrapure water of the product was used as template for the second reaction. PCR products of the second reaction were purified using the Qiagen MinElute PCR Purification Kit (Qiagen, Hilden, Germany). Ultrapure water was used in place of template DNA and run along with each batch of PCRs to serve as a negative control for PCR; none of these produced visible bands on an agarose gel. In total, four separate replicates from each of 31 DNA samples were carried through the two-step PCR process for a total of 124 sequenced PCR products. These were combined with additional samples from other projects, totaling 345 samples for sequencing.

DNA Sequencing

Up to 30 PCR products were combined according to their primer index in equal concentration into one of 14 pools, and 150 nanograms from each were prepared for library sequencing using the KAPA high-throughput library prep kit with real-time library amplification protocol (KAPA Biosystems, Wilmington, MA, USA). Each of these ligated sequencing adapters included an additional 6 base pair index sequence (NEXTflex DNA barcodes; BIOO Scientific, Austin, TX, USA). Thus, each

202 PCR product was identifiable via its unique combination of index sequences in the sequencing
203 adapters and primers. Fragment size distribution and concentration of each library was quantified
204 using an Agilent 2100 BioAnalyzer. Libraries were pooled in equal concentrations and sequenced
205 for 150 base pairs in both directions (PE150) using an Illumina NextSeq at the Stanford Functional
206 Genomics Facility, where 20% PhiX Control v3 was added to act as a sequencing control and to
207 enhance sequencing depth by increasing sequence diversity. Raw sequence data in fastq format is
208 publicly available (see Data Availability).

209 **Sequence Data Processing (Bioinformatics)**

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

219 **Ecological Analyses**

220 After gathering the data, we use the eDNA community observed at each location to make inferences
221 about the spatial patterning of eDNA communities. We use statistical tools from community ecology
222 to assess the spatial structure of eDNA communities. We report similarity (1- dissimilarity) rather
223 than dissimilarity in all cases for ease of interpretation.

224 **Objective 1: Community similarity as a function of distance**

225 **Distance Decay**

226 To address our first objective and determine whether or not nearby samples are more similar than
227 distant ones, we fit a nonlinear model to represent decreasing community similarity with distance.

228 We calculated the pairwise Bray-Curtis similarity (1 - Bray-Curtis dissimilarity) between eDNA
 229 communities using the R package *vegan* (Oksanen et al., 2016) and the great circle distance between
 230 sampling points using the Haversine method as implemented by the R package *geosphere* (Hijmans,
 231 2016). This model is similar to the Michaelis-Menten function, but with an asymptote fixed at 0:

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

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

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

246 **Objective 2: Spatial distribution of diversity**

247 **Community Classification**

248 To determine the spatial distribution and variation of eDNA communities (objective 2), we used
 249 multivariate classification algorithms. We simultaneously assessed the existence of distinct com-
 250 munity types and the membership of samples to those community types using an unsupervised
 251 classification algorithm known as partitioning around medoids (PAM; sometimes referred to as k-
 252 medoids clustering) (Kaufman and Rousseeuw, 1990), as implemented in the R package *cluster*
 253 (Maechler et al., 2016). The classification of samples to communities was made on the basis of

254 their pairwise Bray-Curtis similarity, calculated using the function `vegdist` in the R package `vegan`
 255 (Oksanen et al., 2016). Other distance metrics were evaluated but had no appreciable effect on the
 256 outcome of the analysis (Figure 8). In order to chose an optimal number of clusters (K), we evalu-
 257 ated the distribution of silhouette widths, a measure of the similarity between each sample and its
 258 cluster compared to its similarity to other clusters. We repeated the analysis using fuzzy clustering
 259 (FANNY, (Kaufman and Rousseeuw, 1990); however, the results were qualitatively similar to the
 260 results using PAM so we omit them here.

261 **Aggregate Measures of Diversity**

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

273 **Taxon and Life History Patterns**

274 After assigning taxon names to the abundance data, we plotted the distribution in space of a
 275 selection of taxa to compare with our expectations on the basis of adult distributions (objective 2).
 276 Our aim was to understand where each taxon occurred in the greatest proportional abundance, and
 277 its distribution in space relative to that maximum. Thus, we rescaled each sample to proportional
 278 abundance, extracted the data from a single taxon, and scaled those values between 0 and 1.
 279 We collated life history characteristics for each of the major taxonomic groups recovered, including
 280 dispersal range of the gametes, larvae, and adults, adult habitat type and selectivity, and adult body
 281 size. For each life history stage of each taxon group, we made an order-of-magnitude approximation

282 of the scale of dispersal. For example, internally fertilized species were assigned a gamete range
283 of 0 km, while broadcast spawners were assigned a gamete range of 10 km. Similarly, adult range
284 size was approximated as 0 km (sessile), 1 km (motile but not pelagic), or 10 km (highly mobile,
285 pelagic). Variables were specified as 'multiple' for life history stages known to span more than 1
286 magnitude of range size. For groups to which sequences were annotated with high confidence, but
287 for which life history strategy is diverse or poorly known (e.g. families in the phylum Nemertea),
288 we used conservative, coarse approximations at a higher taxonomic rank (see Life History Data).
289 These data were used to contextualize group-specific spatial distributions and inform expectations
290 based on known adult distributions.

291 **Results**

292 **Sequence Data Processing (Bioinformatics)**

293 Preliminary sequence analysis strongly suggested that the observed variation among environmental
294 samples reflects true variation in the environment, rather than variability due to lab protocols, for
295 the following reasons (note that all value ranges are reported as mean plus and minus one standard
296 deviation). First, all libraries passed the FastQC per-base sequence quality filter, generating a total
297 of 371,576,190 reads passing filter generated in each direction. Second, samples in this study were
298 represented by an adequate number of reads ($333,537.9 \pm 112,200.5$), with no individual sample
299 receiving fewer than 130,402 reads. Third, there was a very low frequency of cross-contamination
300 from other libraries into those reported here ($5e-05 \pm 8e-05$; max proportion 0.00034). Fourth, after
301 scaling all samples to the same sequencing depth, OTUs with abundance greater than 178 reads
302 (0.14% of a sample's reads) experienced no turnover among PCR replicates within a sample. Fifth,
303 sequence abundances among PCR replicates within water samples were remarkably consistent. A
304 single sample had low similarity among PCR replicates (0.659) after removing this outlier, the
305 lowest mean similarity among replicates within a sample was 0.966. Overall similarities among
306 PCR replicates within a sample were extremely high (0.976 ± 0.013), and far higher than those
307 among samples (0.3 ± 0.16). Across PCR replicates, each sample was represented by at least 781425
308 reads in the raw data and contained between 111 and 443 rarefied OTUs (Supplemental Figure 10).

309 **Ecological Analyses**

310 **Distance Decay**

311 Physical proximity is a good predictor of eDNA community similarity: Similarity decreased from
312 0.40 (95%CI = 0.36, 0.45) to half that amount at 4500 meters (95%CI = 2900, 7500) (Figure 2).

313 **Community Classification**

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

325 **Diversity in Space**

326 Sites offshore tend to be less rich and more even than those inshore (Figure 6). Mean OTU richness
327 declined by 1.42 per 1000 meters from a mean of 17.6 taxa (95%CI = 2.15) inshore to 11.9 taxa
328 (95%CI = 4.31) at offshore locations ($p = 0.0415$; Figure 6). Evenness increased by .0666 per 1000
329 meters from 0.225 (95%CI = 0.0558) to 0.491 (95%CI = ± 0.112), indicating that sequence reads
330 were less evenly distributed among taxa in offshore samples ($p \ll 0.05$; Figure 6). There was no
331 evidence for spatial autocorrelation for any of the diversity metrics (Moran's I, $p > 0.05$; Figure 5).

332 **Taxon and Life History Patterns**

333 We were able to assign a taxon name with confidence to 136 of 146 OTU sequences. The vast ma-
334 jority of sequences (97.6%) and OTUs (96.9%) were matched to organisms that have high potential

for dispersal at either the gamete, larval, or adult stage, making it impossible to determine whether the source of that DNA was adults with well-documented spatial patterns (e.g. sessile nearshore specialists) or highly mobile early life history stages. Of the 6 OTUs for which dispersal is limited during all life history stages, only 2 occurred in more than two samples, precluding a quantitative comparison of spatial dispersion based on life history characteristics. These were assigned to *Cymatogaster aggregata*, a viviparous nearshore fish with internal fertilization, and *Cupolaconcha meroclista*, a sessile Vermetid gastropod with presumed internal fertilization and short larval dispersal (Strathmann and Strathmann, 2006; Phillips and Shima, 2010; Calvo and Templado, 2004). *Cymatogaster aggregata* was distinctly more abundant close to shore, with no sequences occurring in any sample beyond 250 meters (Figure 7). *Cupolaconcha meroclista* showed no such distinct spatial trend, occurring in nearly equal abundance at three sites, 75, 500, and 2000 meters from shore. An additional species that was highly abundant in the sequence data, the krill *Thysanoessa raschii*, has pelagic adults, highly seasonal reproduction, and sinking eggs; their distribution was consistent with our expectations based on a tendency of adults to aggregate offshore. Finally, the two most abundant taxa in the dataset were the mussel genus *Mytilus* and the Barnacle order Sessilia; the adults of both taxa are sessile and occur exclusively on hard intertidal substrata but have highly motile larvae.

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. 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 distant locations have less-similar eDNA communities than 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

389 strong tidal exchange in a steep bathymetric setting (Yang and Khangaonkar, 2010). It would be
390 useful to better understand such features during the period of sampling, by way of oceanographic
391 monitoring devices. Finally, the uncertainty in identification of the number of distinct clusters to
392 best characterize the community underlines the difficulty of identifying community patterns with
393 the number of taxonomic groups considered here. We suspect that the signature of eDNA from
394 microscopic life-stages may explain our inability to easily detect spatial community level patterns
395 that align with our initial expectations.

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

397 We found that richness declined while evenness increased with distance from shore. Such a pat-
398 tern is consistent with many other ecosystems which show strong clines in diversity metrics over
399 environmental gradients. However, our study is novel in that it corroborates a cline well-known
400 on macroscales for macrobiota on a much smaller spatial scale for microscopic animals, suggesting
401 that there may be a self-similarity across scales in diversity patterning (Levin, 1992). The coastal
402 ocean is a highly productive and diverse ecosystem, where biomass is concentrated most heavily
403 along the bottom and shoreline (Ray, 1988). This differential in biomass concentration from the
404 shoreline to open waters may contribute to the opposing trends we detected. Where particles (or-
405 ganisms, tissues, and cells) are sparse, fewer would be collected per sample of constant volume, thus
406 decreasing the probability of drawing as many types (richness) and increasing the probability that
407 any two particles originate from the same type (evenness). Intriguingly, the cline in diversity from
408 inshore to offshore was not determined by shared changes in communities as one moved offshore; the
409 classification analysis suggested a fair amount of differences among communities at a given offshore
410 distance (Figure 3).

411 **(4) Spatial patterning of eDNA is associated with taxon-specific life history character-** 412 **istics.**

413 In contrast to our expectations, other taxa including species with sessile adult stages restricted
414 to benthic hard substrates (e.g. barnacles, mussels) are among the most abundant taxa at sites
415 furthest from shore. However, the larvae and gametes of these taxa are abundant, pelagic, and
416 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 than eDNA, affecting the observed community. Our sampling bottles excluded particles larger than 500 micrometers, but gametes and very small larvae could have gained entry. It is possible that even a single small individual, containing many thousand mitochondria, would overwhelm the signal of another species from which hundreds of cells had been sloughed from many, larger individuals. Data on larval size distribution at the time of sampling from each species in our data set would allow us to estimate the frequency of such events. Nevertheless, it is precisely the sensitivity to small particles that makes the eDNA approach powerful, so we are reluctant to recommend that aquatic eDNA sampling use finer pre-filtering. Instead, we emphasize the importance of designing and selecting primer sets that selectively amplify target organisms. In the case of the present study, in order to recover patterns matching our expectations, this would be non-transient, benthic marine organisms lacking any pelagic life stage.

The marker we chose for this study detects a wide variety of metazoans while excluding other more common taxa; however, it does not effectively discriminate among species within a higher group in all cases. Other markers, such as mitochondrial cytochrome c oxidase subunit 1 (COX1, CO1, or COI) may provide adequate species-level resolution in some metazoan groups, but have other shortcomings including taxon dropout (Deagle et al., 2014) and amplification of more abundant non-metazoans, as we discovered in an accompanying study (Kelly et al., 2016a). Both have undesirable effects of biasing estimates of diversity. In our case, it is possible that the lumping of multiple species into one group underestimates the true richness of the group and of the entire sample, in

turn obscuring true underlying patterns of diversity. In the case of COX1, well-documented primer biases cause failure to amplify some taxa, particularly in mixed samples, with the same result (Deagle et al., 2014). In fact, even surveys relying on traditional capture techniques (e.g. seine nets) and morphological characteristics are subject to biases imposed by the sampling gear (e.g. mesh size), the observer (e.g. taxonomic expertise), and organisms (e.g. morphologically cryptic species). Similarly, no single molecular marker adequately and effectively samples all taxa without bias (Drummond et al., 2015), and thus the choice of marker is an important and context-dependent one. Until whole-genome sequencing of individual cells is a reality, the tradeoffs between taxonomic breadth and resolution will continue to be problematic for metabarcoding studies, just as they are for more traditional ecological survey methods (Kelly et al., 2016a).

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

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

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

481 **Acknowledgements**

482 We wish to thank Linda Park, Robert Morris, E. Virginia Armbrust, and James Kralj. The
483 manuscript was improved by suggestions from editor Magnus Johnson, and reviewers Owen Wan-
484 gensteen and Stephen Moss.

485 **Funding**

486 This work was supported by a grant from the David and Lucile Packard Foundation to RPK (grant
487 2014-39827). The funders had no role in study design, data collection and analysis, decision to
488 publish, or preparation of the manuscript.

489 **Author Contributions**

490 Conceived and designed the experiments: JL O'Donnell, RP Kelly, AO Shelton; Collected the data:
491 JL O'Donnell, NC Lowell, GD Williams, RP Kelly, AO Shelton, JF Samhouri; Conducted the
492 analyses: JL O'Donnell; Wrote the first draft: JL O'Donnell; Edited the manuscript: JL O'Donnell,
493 AO Shelton, RP Kelly, JF Samhouri, GD Williams, NC Lowell

494 **Ethics Statement**

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

498 **Data Availability**

499 **Sequence Data**

500 All sequence data are available from NCBI under BioProject PRJNA338801.

501 Scripts to process raw sequence data into the contingency tables used for ecological analyses can be
502 found at:

503 <https://github.com/jimmyodonnell/banzai>

504

505 **Project Repository**

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

507 https://github.com/jimmyodonnell/Carkeek_eDNA_grid

508 <http://dx.doi.org/FIXME>

509 **Sequencing Metadata**

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

511 **Life History Data**

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

513 **Analysis Scripts**

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

516 **References**

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

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

524 Barnes, M. A., Turner, C. R., Jerde, C. L., Renshaw, M. A., Chadderton, W. L., and Lodge, D. M.
 525 (2014). Environmental conditions influence eDNA persistence in aquatic systems. *Environmental*
 526 *Science and Technology*, 48(3):1819–1827.

527 Bell, T. (2010). Experimental tests of the bacterial distance–decay relationship. *The ISME Journal*,
 528 4(11):1357–1365.

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

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

532 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L.
 533 (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10:421.

534 Chamberlain, S. a. and Szöcs, E. (2013). taxize: taxonomic search and retrieval in R. *F1000Research*,
 535 2(0):191.

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

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

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

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

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

548 Cowart, D. a., Pinheiro, M., Mouchel, O., Maguer, M., Grall, J., Miné, J., and Arnaud-Haond, S.
549 (2015). Metabarcoding Is Powerful yet Still Blind: A Comparative Analysis of Morphological and
550 Molecular Surveys of Seagrass Communities. *Plos One*, 10(2):e0117562.

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

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

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

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

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

572 Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*,
573 26(19):2460–2461.

574 Evans, N. T., Olds, B. P., Renshaw, M. A., Turner, C. R., Li, Y., Jerde, C. L., Mahon, A. R.,
575 Pfreder, M. E., Lamberti, G. A., and Lodge, D. M. (2016). Quantification of mesocosm fish and
576 amphibian species diversity via environmental DNA metabarcoding. *Molecular Ecology Resources*,
577 16(1):29–41.

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

580 Fonseca, V. G., Carvalho, G. R., Nichols, B., Quince, C., Johnson, H. F., Neill, S. P., Lamshead,
581 J. D., Thomas, W. K., Power, D. M., and Creer, S. (2014). Metagenetic analysis of patterns of
582 distribution and diversity of marine meiobenthic eukaryotes. *Global Ecology and Biogeography*,
583 23(11):1293–1302.

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

585 Guardiola, M., Uriz, M. J., Taberlet, P., Coissac, E., Wangenstein, O. S., and Turon, X.
586 (2015). Deep-Sea, Deep-Sequencing: Metabarcoding Extracellular DNA from Sediments of Ma-
587 rine Canyons. *PLOS ONE*, 10(10):e0139633.

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

591 Guo, H., Chamberlain, S. A., Elhaik, E., Jalli, I., Lynes, A. R., Marczak, L., Sabath, N., Vargas,
592 A., Więski, K., Zelig, E. M., and Pennings, S. C. (2015). Geographic variation in plant commu-
593 nity structure of salt marshes: Species, functional and phylogenetic perspectives. *PLoS ONE*,
594 10(5):e0127781.

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

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

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

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

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

605 Kelly, R. P., Closek, C. J., O’Donnell, J. L., Kralj, J. E., Shelton, A. O., and Samhour, J. F.
 606 (2016a). Genetic and manual survey methods yield different and complementary views of an
 607 ecosystem. *Frontiers in Marine Science*, 3:283.

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

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

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

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

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

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

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

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

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

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

630 Martiny, J. B. H., Eisen, J. A., Penn, K., Allison, S. D., and Horner-Devine, M. C. (2011). Drivers of
 631 bacterial beta diversity depend on spatial scale. *Proceedings of the National Academy of Sciences*,
 632 108(19):7850–7854.

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

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

637 Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R.,
 638 O’Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., and Wagner, H.
 639 (2016). *vegan: Community Ecology Package*.

640 Parr, C. S., Wilson, N., Leary, P., Schulz, K. S., Lans, K., Walley, L., Hammock, J. A., Goddard,
 641 A., Rice, J., Studer, M., Holmes, J. T. G., and Corrigan, R. J. (2014). The Encyclopedia of
 642 Life v2: Providing Global Access to Knowledge About Life on Earth. *Biodiversity Data Journal*,
 643 2(2):e1079.

644 Phillips, N. E. and Shima, J. S. (2010). Reproduction of the vermetid gastropod *dendropoma*
 645 *maximum* (Sowerby, 1825) in Moorea, French Polynesia. *Journal of Molluscan Studies*, 76(2):133–
 646 137.

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

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

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

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

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

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

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

664 Shelton, A. O., O'Donnell, J. L., Samhour, J. F., Lowell, N., Williams, G. D., and Kelly, R. P.
665 (2016). A framework for inferring biological communities from environmental DNA. *Ecological*
666 *Applications*, 26(6):1645–1659.

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

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

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

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

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

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

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

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

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

687 Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M.,
688 Solovyev, V. V., Rubin, E. M., Rokhsar, D. S., and Banfield, J. F. (2004). Community structure
689 and metabolism through reconstruction of microbial genomes from the environment. *Nature*,
690 428(6978):37–43.

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

696 Wetzel, C. E., de Bicudo, D. C., Ector, L., Lobo, E. A., Soininen, J., Landeiro, V. L., and Bini,
697 L. M. (2012). Distance Decay of Similarity in Neotropical Diatom Communities. *PLoS ONE*,
698 7(9):e45071.

699 Yang, Z. and Khangaonkar, T. (2010). Multi-scale modeling of Puget Sound using an unstructured-
700 grid coastal ocean model: From tide flats to estuaries and coastal waters. *Ocean Dynamics*,
701 60(6):1621–1637.

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

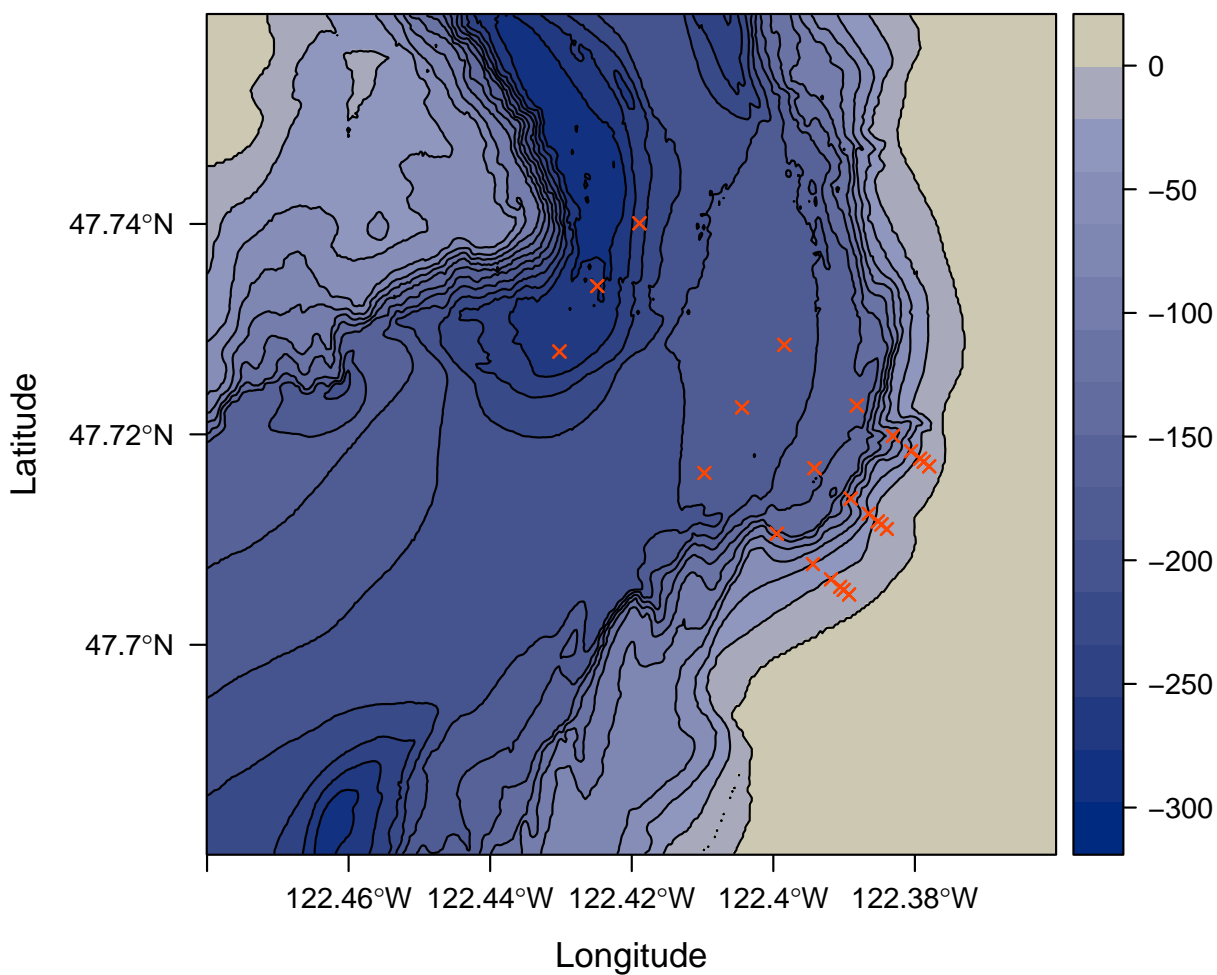


Figure 1: Map of study area. Depth in meters below sea level is indicated by shading and 25 meter contours. Sampled locations are indicated by red points.

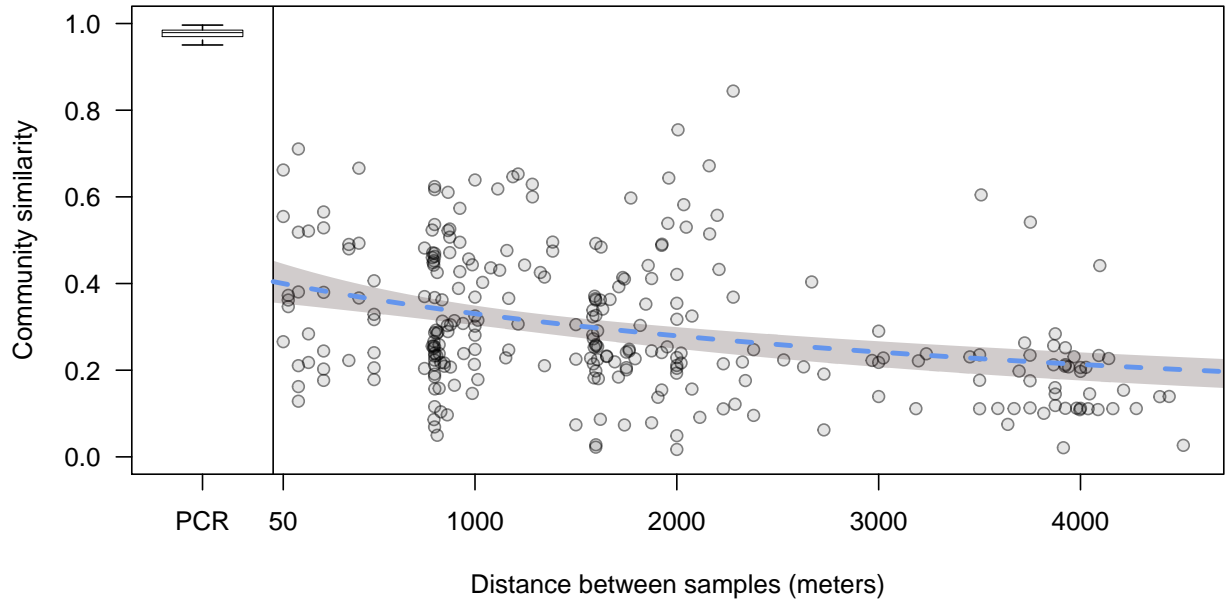


Figure 2: Distance decay relationship of environmental DNA communities. Each point represents the Bray-Curtis similarity of a site sampled along three parallel transects comprising a 3000 by 4000 meter grid. Blue dashed line represents fit of a nonlinear least squares regression (see Methods), and shading 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.

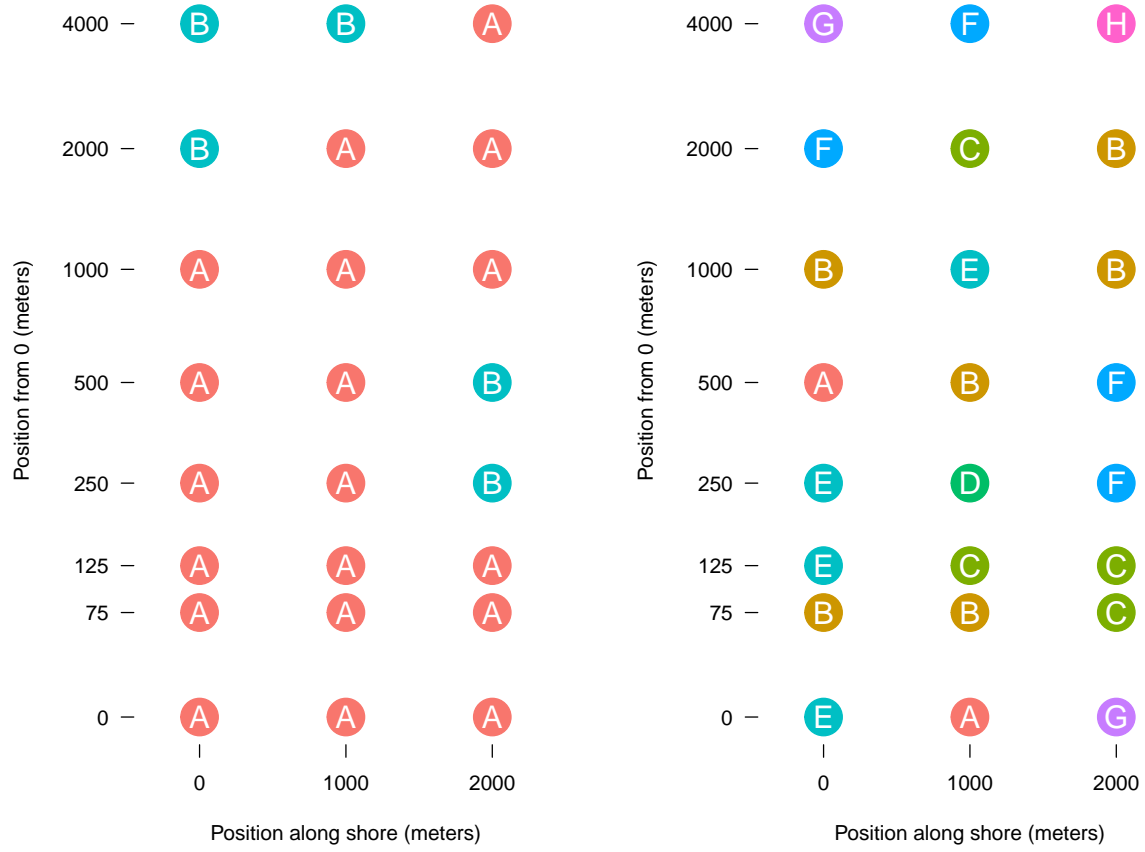


Figure 3: Cluster membership of sampled sites. Distance from onshore starting point is log scaled. Sites are colored and labeled by their assignment to a cluster by PAM analysis for number of clusters (K) chosen based on a priori expectations (2) and mean silhouette width (8).

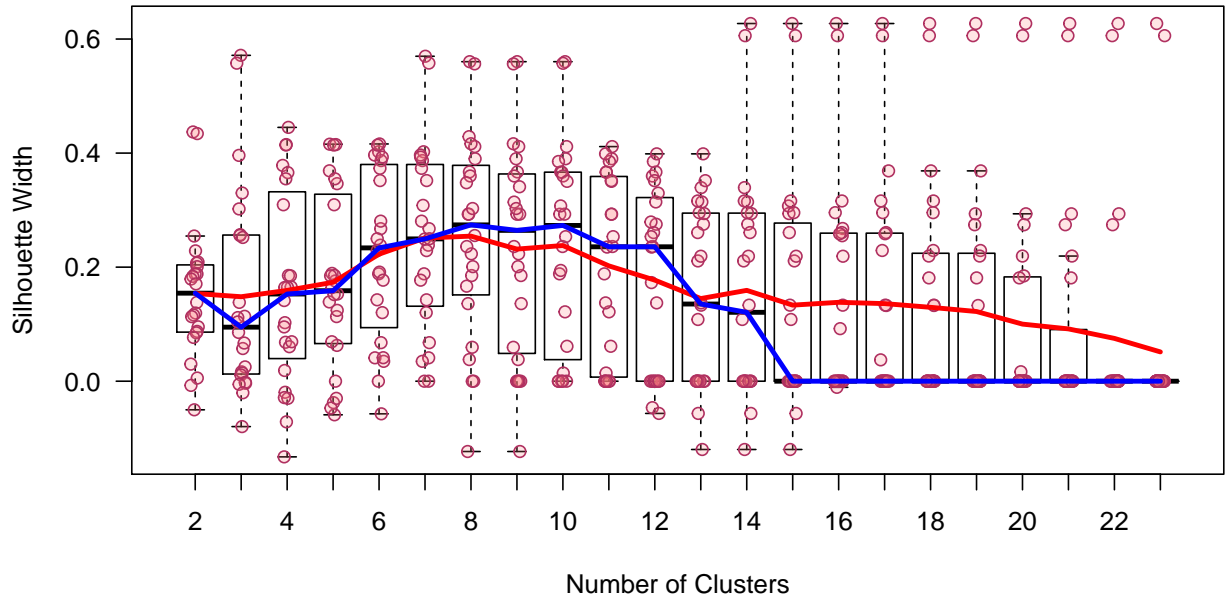


Figure 4: Silhouette widths from PAM analysis. Points are the width of the PAM silhouette of each sample at each number of clusters (K). Red line is the mean, blue line is the median. Boxes encompass the interquartile range with a line at the median, and the whiskers extend to 1.5 times the interquartile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

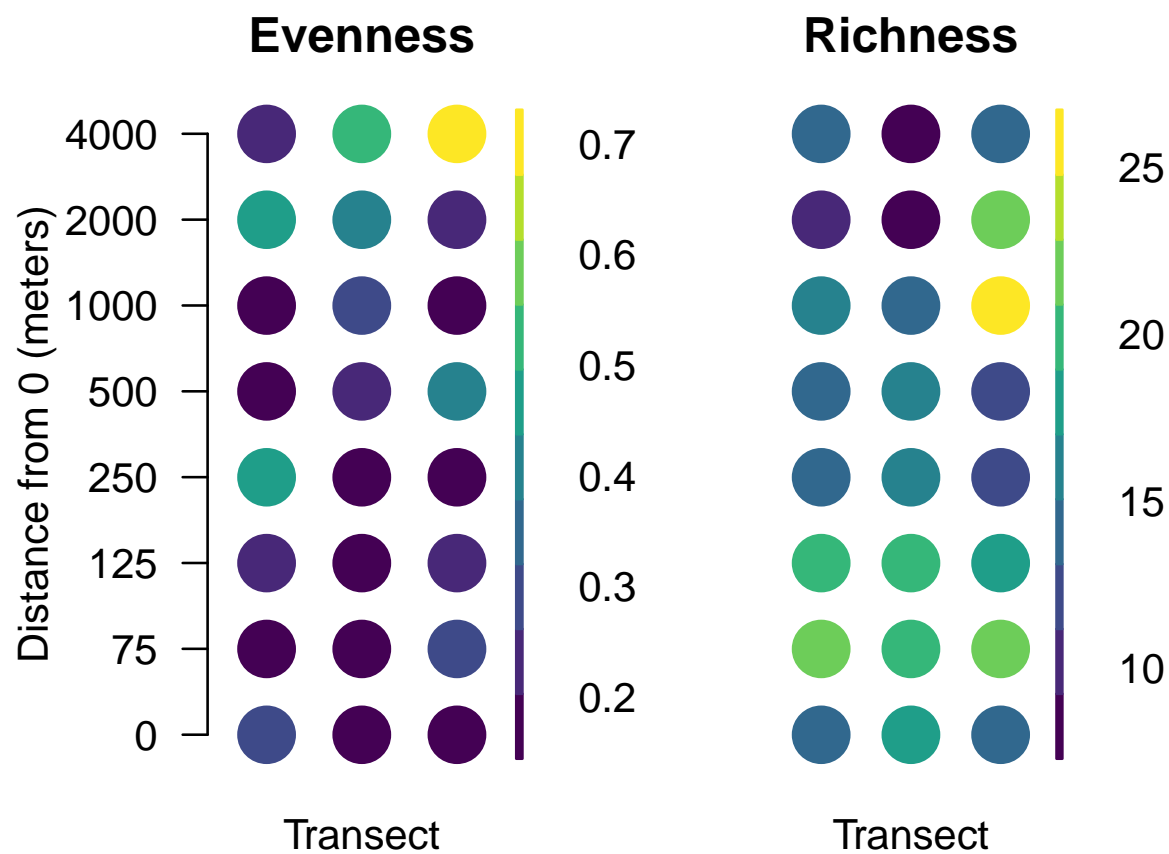


Figure 5: Aggregate measures of diversity at each sample site. Data are rarefied counts of mitochondrial 16S sequences collected from 3 parallel transects in Puget Sound, Washington, USA. Evenness (left) is the probability that two sequences drawn at random are different; richness (right) represents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

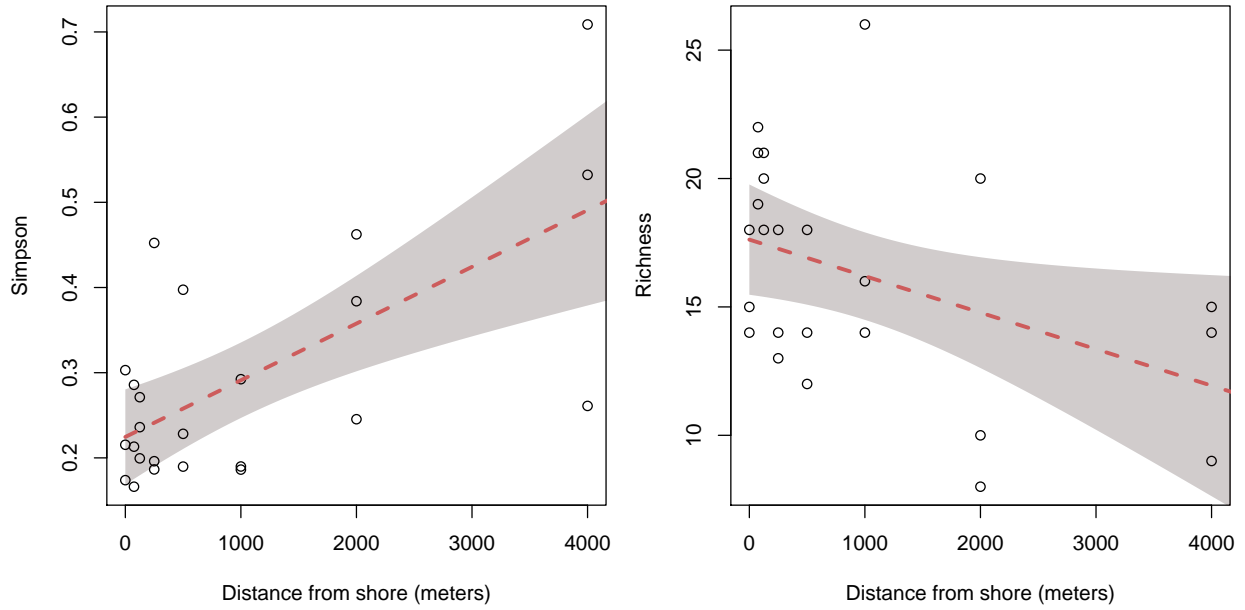


Figure 6: Aggregate diversity metrics of each site plotted against distance from shore. Both Simpson's Index (left) and richness (right) are shown, and have been computed from the mean abundance of unique DNA sequences found across 4 PCR replicates at each of 24 sites. Lines and bands illustrate the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

705 **Supplemental Material**

706 **Methods**

707 **Bioinformatics**

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

717 We used vsearch v2.1.1 (Rognes et al., 2016) to discard any merged reads for which the sum of the
718 per-base error probabilities was greater than 0.5 (“expected errors”) Edgar (2010). Sequences were
719 demultiplexed on the basis of the primer index sequence at base positions 4-9 at both ends using the
720 programming language AWK. Primer sequences were removed using cutadapt v1.7.1 Martin (2011),
721 allowing for 2 mismatches in the primer sequence. Identical duplicate sequences were identified,
722 counted, and removed in python to speed up subsequent steps by eliminating redundancy, and
723 sequences occurring only once were removed. We checked for and removed any sequence likely to be
724 a PCR artifact due to incomplete extension and subsequent mis-priming using a method described
725 by Edgar (2010) and implemented in vsearch v2.0.2. Sequences were clustered into operational
726 taxonomic units (OTUs) using the single-linkage clustering method implemented by swarm version
727 2.1.1 with a local clustering threshold (d) of 1 and fastidious processing (Mahé et al., 2014).

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

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

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

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

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

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

763 tive sequence from each OTU was then queried against the NCBI nucleotide collection (GenBank;
764 version October 7, 2015; 32,827,936 sequences) using the blastn command line utility (Camacho
765 et al., 2009). In order to maximize the accuracy of this computationally intensive step, we imple-
766 mented a nested approach whereby each sequence was first queried using strict parameters (e-value
767 = 5e-52), and if no match was found, the query was repeated with decreasingly strict e-values (5e-48
768 5e-44 5e-40 5e-36 5e-33 5e-29 5e-25 5e-21 5e-17 5e-13). Other parameters were unchanged among
769 repetitions (word size: 7; maximum matches: 1000; culling limit: 100; minimum percent identity:
770 0). Each query sequence can be an equally good match to multiple taxa either because of invariabil-
771 ity among taxa or errors in the database (e.g. human sequences are commonly attributed to other
772 organisms when they in fact represent lab contamination). In order to guard against these spurious
773 results, we used an algorithm to find the lowest common taxon for at least 80% of the matched
774 taxa, implemented in the R package taxize 0.7.8 (Chamberlain and Szöcs, 2013; Chamberlain et al.,
775 2016). Similarly, we repeated analyses using the dataset consolidated at the same taxonomic rank
776 across all queries, for the rank of both family and order.

777 **Alternative distance decay model formulations**

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

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

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

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

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

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

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

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

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

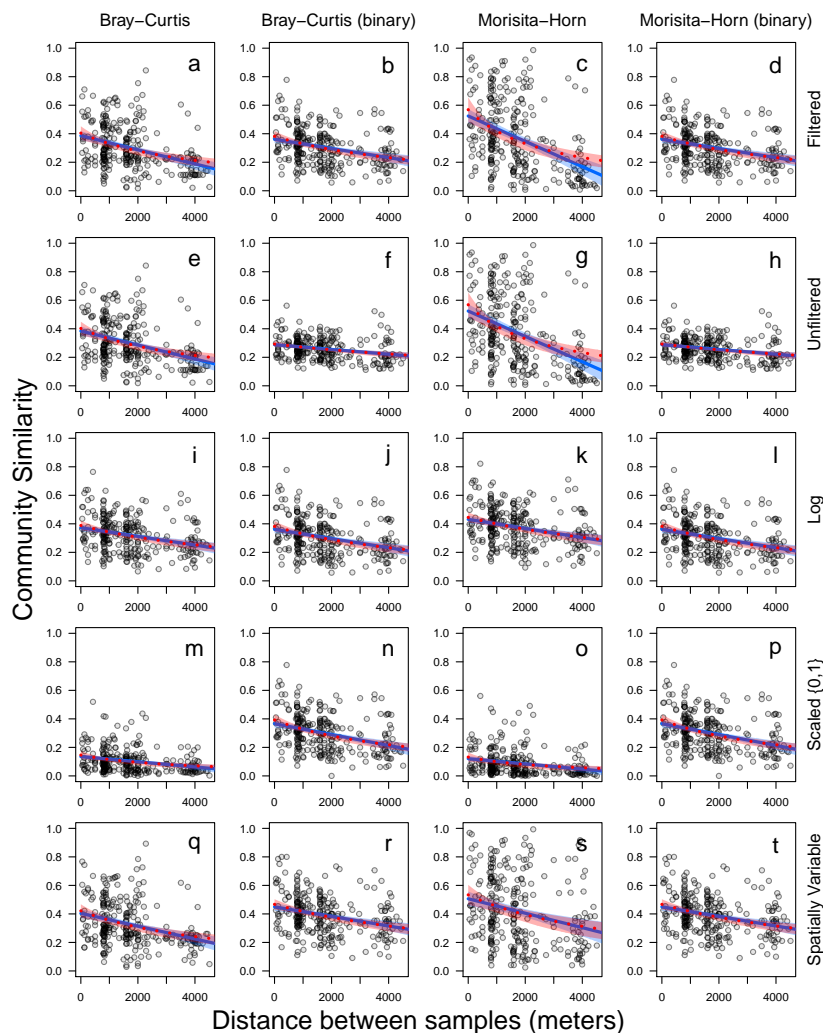


Figure 8: Distance decay relationship of environmental DNA communities using a variety of models, metrics, and data subsets. Each point represents the similarity of a site sampled along three parallel transects comprising a 3000 by 4000 meter grid. Each row of plots represents a different data subset 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.

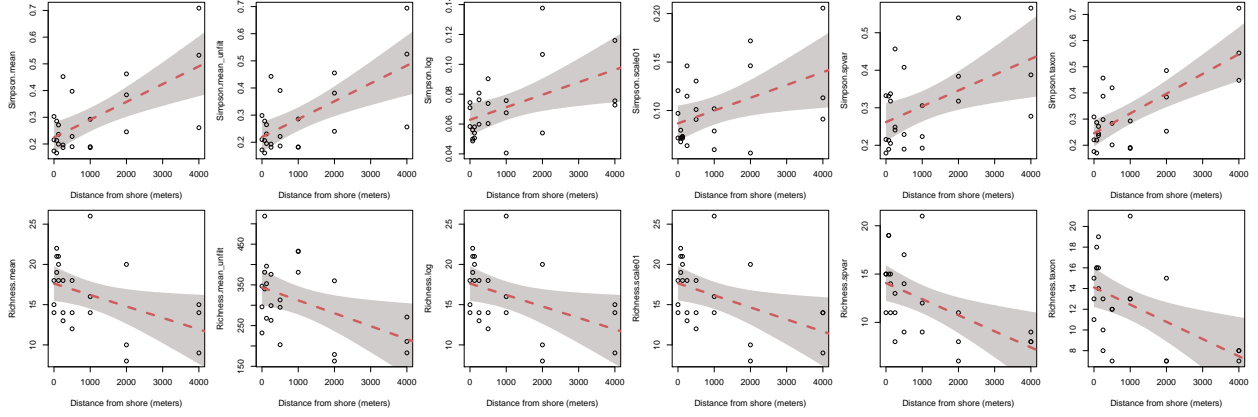


Figure 9: Aggregate diversity metrics of each site plotted against distance from shore. Both Simpson's Index (top) and richness (bottom) are shown for a variety of data subsets and transformations (left to right: mean, unfiltered mean, $\log(x + 1)$, transformed, scaled, spatially variable, and taxon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

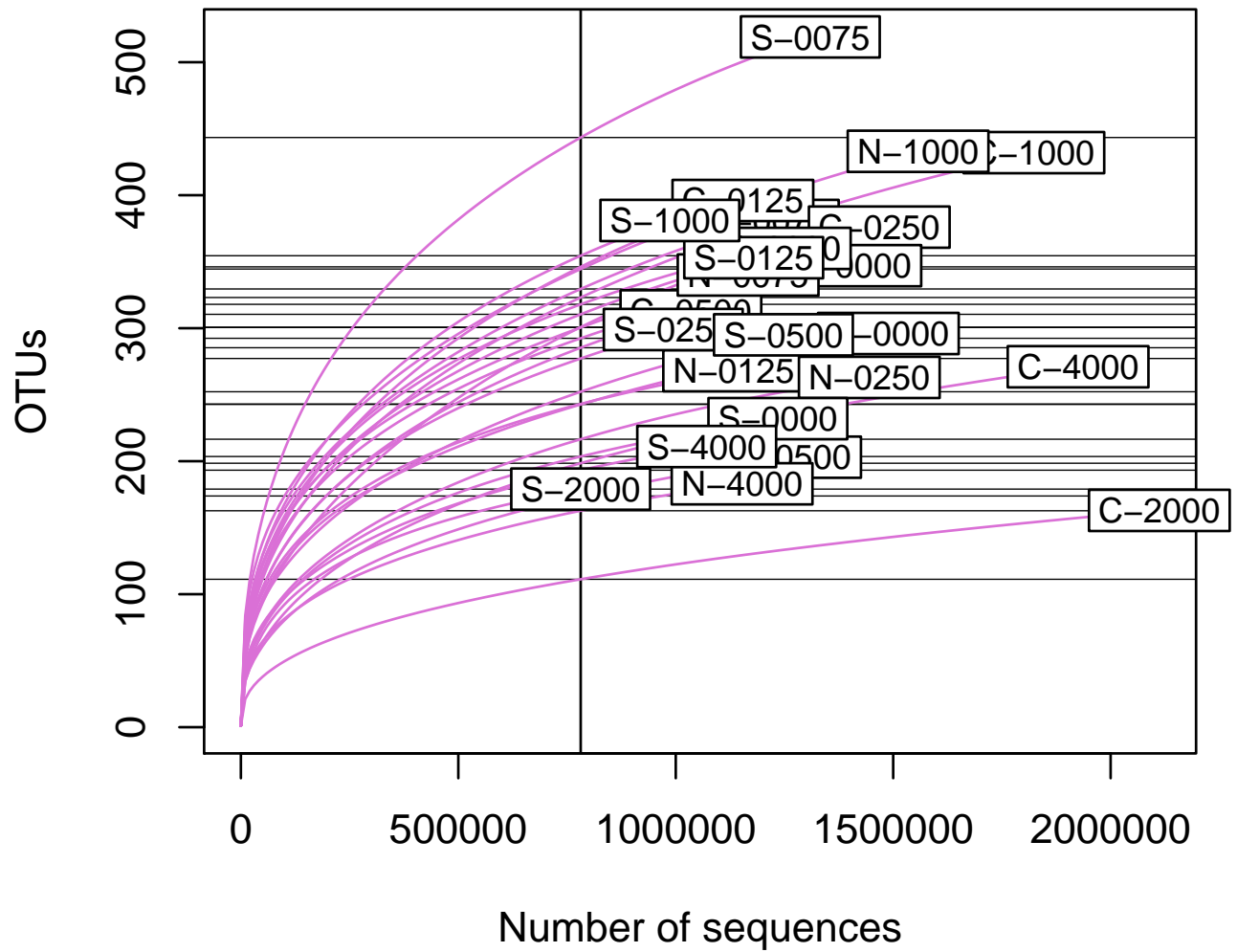


Figure 10: Accumulation of OTUs from 24 environmental samples using randomized rarefaction. Four replicate PCRs were conducted using DNA each environmental sample and independently sequenced, but these are collapsed here to illustrate a single representation of richness. Sample names indicate the position 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.