Spatial distribution of environmental DNA in a nearshore marine habitat

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

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

[2]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

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

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

[5]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 19, 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 (Cowart 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

59 a key step in making these methods practical.

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

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

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

88   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, copepods, 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

119 diversity measures of eDNA communities that have been widely used to reveal broad scale patterns

120 based on macrobiota in many ecological systems. Finally, we identify the taxa represented in the

121 eDNA communities, which span a range of life-history characteristics, and we expect that the spatial

122 distribution of eDNA will most closely resemble the distribution of adults in taxa with low dispersal

123 potential.

Methods

125 There are seven discrete steps to our methodology: (1) Environmental sample collection, (2) isolation

126 of particulates from water via filtration, (3) isolation of DNA from filter membrane, (4) amplification

127 of target locus via PCR, (5) sequencing of amplicons, (6) bioinformatic translation of raw sequence

128 data into tables of sequence abundance among samples, and (7) community ecological analyses of

129 eDNA. We provide brief overviews of these steps here, and encourage the reader to review the fully

130 detailed methods presented in the supplementary material.

Environmental Sampling

132 Starting from lower-intertidal patches of *Zostera marina*, we collected water samples at 1 meter

133 depth from 8 points (0, 75, 125, 250, 500, 1000, 2000, and 4000 meters) along three parallel transects

134 separated by 1000 meters (24 sample locations total; Figure 1). Samples were collected by attaching

135 bottles to a PVC pole and lowering it over the side of a boat over the span of one hour on 27 June

136 2014. To destroy residual DNA on equipment used for field sampling and filtration, we washed

137 with a 1:10 solution of household bleach (8.25% sodium hypochlorite; 7.25% available chlorine) and

138 deionized water, followed by thorough rinsing with deionized water. Each environmental sample

139 was collected in a clean 1 liter high-density polyethylene bottle, the opening of which was covered

140 with 500 micrometer nylon mesh to prevent entry of larger particles. Immediately after collecting

141 the sample, the mesh was replaced with a clean lid and the sample was held on ice until filtering.

Filtration

143 One liter from each water sample was filtered in the lab on a clean polysulfone vacuum filter

144 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.

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.

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

173 index sequences differing by a minimum Hamming distance of 3 (see Sequencing Metadata). In-

174 dexed primers were assigned to samples randomly, with the identical index sequence on the forward

175 and reverse primer to avoid errors associated with dual-indexed multiplexing (Schnell et al., 2015).

176 In a UV-sterilized hood, we prepared 25 microliter reactions containing 18.375 microliters ultrapure

177 water, 2.5 microliters 10x buffer, 0.625 microliters deoxynucleotide solution (8 millimolar), 1 micro-

178 liter each forward and reverse primer (10 micromolar, obtained lyophilized from Integrated DNA

179 Technologies (Coralville, IA, USA)), 0.25 microliters Qiagen HotStar Taq polymerase, and 1.25

180 microliter genomic or eDNA template at 1:100 dilution in ultrapure water. PCR thermal profiles

181 began with an initialization step (95 ºC; 15 min) followed by cycles (40 and 20 for the first and

182 second reaction, respectively) of denaturation (95 ºC; 15 sec), annealing (61 ºC; 30 sec), and exten-

183 sion (72 ºC; 30 sec). 20 identical PCRs were conducted from each DNA extract using non-indexed

184 primers; these were pooled into 4 groups of 5 in order to ensure ample template for the subsequent

185 PCR with indexed primers. In order to isolate the fragment of interest from primer dimer and

186 other spurious fragments generated in the first PCR, we used the AxyPrep Mag FragmentSelect-I

187 kit with solid-phase reversible immobilization (SPRI) paramagnetic beads at 2.5x the volume of

188 PCR product (Axygen BioSciences, Corning, NY, USA). A 1:5 dilution in ultrapure water of the

189 product was used as template for the second reaction. PCR products of the second reaction were

190 purified using the Qiagen MinElute PCR Purification Kit (Qiagen, Hilden, Germany). Ultrapure

191 water was used in place of template DNA and run along with each batch of PCRs to serve as a

192 negative control for PCR; none of these produced visible bands on an agarose gel. In total, four

193 separate replicates from each of 31 DNA samples were carried through the two-step PCR process

194 for a total of 124 sequenced PCR products. These were combined with additional samples from

195 other projects, totaling 345 samples for sequencing.

196 **DNA Sequencing**

197 Up to 30 PCR products were combined according to their primer index in equal concentration into

198 one of 14 pools, and 150 nanograms from each were prepared for library sequencing using the KAPA

199 high-throughput library prep kit with real-time library amplification protocol (KAPA Biosystems,

200 Wilmington, MA, USA). Each of these ligated sequencing adapters included an additional 6 base

201 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).

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 reads

213 (Zhang et al., 2014), (2) eliminate low-quality reads (Edgar, 2010; Rognes et al., 2016), (3) eliminate

214 PCR artifacts (chimeras) (Edgar, 2010; Rognes et al., 2016; Martin, 2011), (4) cluster reads by

215 similarity into operational taxonomic units (OTUs) (Mahé et al., 2014), and (5) match observed

216 sequences to taxon names (Camacho et al., 2009; Chamberlain and Szöcs, 2013; Chamberlain et al.,

217 2016). Additionally, we checked for consistency among PCR replicates, excluded extremely rare

218 sequences, and rescaled (rarefied) the data to account for differences in sequencing depth. The data

219 for input to further analyses are a contingency table of the mean count of unique sequences, OTUs,

220 or taxa present in each environmental sample.

Ecological Analyses

222 After gathering the data, we use the eDNA community observed at each location to make inferences

223 about the spatial patterning of eDNA communities. We use statistical tools from community ecology

224 to assess the spatial structure of eDNA communities. We report similarity (1- dissimilarity) rather

225 than dissimilarity in all cases for ease of interpretation.

Objective 1: Community similarity as a function of distance

Distance Decay

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

$$y_{ij} = \frac{AB}{B + x_{ij}} \tag{1}$$

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

There are other conceptually reasonable forms to expect the space-by-similarlity relationship to take; we present these in the supplemental material along with alternative data subsets and similarity indices.

Objective 2: Spatial distribution of diversity

Community Classification

To determine the spatial distribution and variation of eDNA communities (objective 2), we used multivariate classification algorithms. We simultaneously assessed the existence of distinct community types and the membership of samples to those community types using an unsupervised classification algorithm known as partitioning around medoids (PAM; sometimes referred to as k-medoids clustering) (Kaufman and Rousseeuw, 1990), as implemented in the R package cluster (Maechler et al., 2016). The classification of samples to communities was made on the basis of their pairwise Bray-Curtis similarity, calculated using the function vegdist in the R package vegan (Oksanen et al., 2016). Other distance metrics were evaluated but had no appreciable effect on the outcome of the analysis (Supplemental Figure 1). In order to chose an optimal number of clusters (K), we evaluated the distribution of silhouette widths, a measure of the similarity between each sample and its cluster compared to its similarity to other clusters. We repeated the analysis using fuzzy clustering (FANNY, (Kaufman and Rousseeuw, 1990); however, the results were qualitatively similar to the results using PAM so we omit them here.

Aggregate Measures of Diversity

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

Taxon and Life History Patterns

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

Results

Sequence Data Processing (Bioinformatics)

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

302 from other libraries into those reported here (5e-05±8e-05; max proportion 0.00034). Fourth, after

303 scaling all samples to the same sequencing depth, OTUs with abundance greater than 178 reads

304 (0.14% of a sample's reads) experienced no turnover among PCR replicates within a sample. Fifth,

305 sequence abundances among PCR replicates within water samples were remarkably consistent. A

306 single sample had low similarity among PCR replicates (0.659) after removing this outlier, the

307 lowest mean similarity among replicates within a sample was 0.966. Overall similarities among

308 PCR replicates within a sample were extremely high (0.976 ± 0.013), and far higher than those

309 among samples (0.3 ± 0.16). Across PCR replicates, each sample was represented by at least 781425

310 reads in the raw data and contained between 111 and 443 rarefied OTUs (Supplemental Figure 2).

311 **Ecological Analyses**

312 **Distance Decay**

313 Physical proximity is a good predictor of eDNA community similarity: Similarity decreased from

314 0.40 (95%CI = 0.36, 0.45) to half that amount at 4500 meters (95%CI = 2900, 7500) (Figure 2).

315 **Community Classification**

316 Despite a clear trend in community similarity as a function of spatial separation, the results from

317 our classification analysis are difficult to interpret. The silhouette analysis indicated the presence

318 of 8 distinct communities; however, the gain in mean silhouette width from 2 was small (0.1), and

319 lacked a distinctive peak (Figure 3), indicating substantial uncertainty in the clustering algorithm.

320 Thus, we present the results of cluster assignment for both $K = 2$ and $K = 8$ to illustrate the

321 range of results (Figure 4). Excluding taxa which occur in only one site had no discernible effect

322 on the outcome of the PAM analysis (number of clusters, assignment to clusters). While there was

323 no distinct spatial divide indicating the presence of an inshore versus an offshore community, one

324 of the two communities (at $K = 2$) occurred in only 2 out of 18 samples inside 1000 meters from

325 shore, and never occurred within 125 meters of shore, suggesting the presence of an inshore and

326 offshore community.

Diversity in Space

Sites offshore tend to be less rich and more even than those inshore (Figure 5). Mean OTU richness declined by 1.42 per 1000 meters from a mean of 17.6 taxa (95%CI = 2.15) inshore to 11.9 taxa (95%CI = 4.31) at offshore locations (p = 0.0415; Figure 5). Evenness increased by .0666 per 1000 meters from 0.225 (95%CI = 0.0558) to 0.491 (95%CI = ± 0.112), indicating that sequence reads were less evenly distributed among taxa in offshore samples (p ≪ 0.05; Figure 5). The subset of data used for this analysis had no qualitative effect on the outcome of this analysis (Supplemental Figure 3). There was no evidence for spatial autocorrelation for any of the diversity metrics (Moran's I, p > 0.05; Figure 6).

Taxon and Life History Patterns

We were able to assign a taxon name with confidence to 136 of 146 OTU sequences. The vast majority 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

355 highly motile larvae. Because large-scale dispersal could not be ruled out for the vast majority of
356 taxa, subsetting the community data by taxonomic group had no qualitative effect on the spatial
357 patterning or diversity metrics, and we omit those results here.

Discussion

359 Indirect surveys of organismal presence are a key development in ecosystem monitoring in the face
360 of increased anthropogenic pressure and dwindling resources for ecological research. Monitoring
361 of organisms using environmental DNA is an especially promising method, given the rapid pace
362 of advancement in technological innovation and cost efficiency in the field of DNA sequencing and
363 quantification. We document four key patterns: (1) eDNA communities far from one another tend to
364 be less similar than those that are nearby, (2) distinct eDNA communities exist and are distributed
365 in a non-random fashion, (3) diversity declines with distance from shore, and (4) spatial patterning
366 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

369 We demonstrate that distant locations have less-similar eDNA communities than proximate loca-
370 tions in Puget Sound, a dynamic marine environment. Our finding is in line with observations
371 based on traditional surveys of terrestrial plants and fungi (Nekola and White, 1999; Bahram et al.,
372 2013; Condit, 2002; Chust et al., 2006) and of microorganisms in freshwater (Wetzel et al., 2012),
373 marine (Chust et al., 2013), and estuarine (Martiny et al., 2011) environments. To our knowledge,
374 it is the first to report such a pattern using massively parallel sequencing of environmental DNA
375 in the marine environment, and the first using any technique to describe this pattern from mac-
376 robial metazoans. We note that the theoretical expectation is that samples at very close distance
377 be nearly completely similar, while our samples separated by the 50 meters were only 40% similar.
378 We interpret this to reflect the highly dynamic nature of this environment, which could cause DNA
379 to be distributed quickly from its source, eroding the rise in similarity at small distances. At the
380 same time, community similarity decreased to very low levels at larger scales, indicating that DNA
381 distribution is not completely unpredictable. This finding implies that the effectively sampled area

382   of individual water samples for eDNA analysis is likely to be quite small (<100m) in this nearshore

383   environment. Our estimated distance-decay relationship does indicate that proximate samples are

384   more similar than distant samples, but we suggest this pattern is partially obscured by other factors,

385   including signal from mobile, microscopic life-stages.

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

387   We demonstrate strong evidence for distinct community types and the non-random spatial pat-

388   terning of those communities. While the spatial distributions of communities is surprising if one

389   were concerned only with the macroscopic life stages of metazoans, it indeed does align with the

390   broader view that even offshore pelagic communities are comprised of and influenced by nearshore

391   organisms. This result underscores the idea that areas immediately offshore act as ecotones, a mix-

392   ing zone of taxa characteristic of benthic and pelagic environments. While there was no distinct

393   break in community types between onshore and offshore sites, there was some clustering of commu-

394   nity types that may be explained by oceanographic features such as nearshore eddies generated by

395   strong tidal exchange in a steep bathymetric setting (Yang and Khangaonkar, 2010). It would be

396   useful to better understand such features during the period of sampling, by way of oceanographic

397   monitoring devices. Finally, the uncertainty in identification of the number of distinct clusters to

398   best characterize the community underlines the difficulty of identifying community patterns with

399   the number of taxonomic groups considered here. We suspect that the signature of eDNA from

400   microscopic life-stages may explain our inability to easily detect spatial community level patterns

401   that align with our initial expectations.

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

403   We found that richness declined while evenness increased with distance from shore. Such a pat-

404   tern is consistent with many other ecosystems which show strong clines in diversity metrics over

405   environmental gradients. However, our study is novel in that it corroborates a cline well-known

406   on macroscales for macrobiota on a much smaller spatial scale for microscopic animals, suggesting

407   that there may be a self-similarity across scales in diversity patterning (Levin, 1992). The coastal

408   ocean is a highly productive and diverse ecosystem, where biomass is concentrated most heavily

409   along the bottom and shoreline (Ray, 1988). This differential in biomass concentration from the

410   shoreline to open waters may contribute to the opposing trends we detected. Where particles (or-

411   ganisms, tissues, and cells) are sparse, fewer would be collected per sample of constant volume, thus

412   decreasing the probability of drawing as many types (richness) and increasing the probability that

413   any two particles originate from the same type (evenness). Intriguingly, the cline in diversity from

414   inshore to offshore was not determined by shared changes in communities as one moved offshore; the

415   classification analysis suggested a fair amount of differences among communities at a given offshore

416   distance (Figure 4).

(4) Spatial patterning of eDNA is associated with taxon-specific life history

418   In contrast to our expectations, other taxa including species with sessile adult stages restricted

419   to benthic hard substrates (e.g. barnacles, mussels) are among the most abundant taxa at sites

420   furthest from shore. However, the larvae and gametes of these taxa are abundant, pelagic, and

421   can be transported long distances by water movement (Strathmann, 1987). This indicates that we

422   likely detected DNA of their pelagic phase gametes and larvae. It is always possible that DNA

423   of adults was advected over long distances and detected offshore but in light of our results with

424   krill and surfperch, we view this as unlikely. We interpret our results as evidence that the chaotic

425   spatial distribution of eDNA communities (Figure 4) results from our primers' affinity for many

426   species which at some point exist as microscopic pelagic gametes or larvae. Our results emphasize

427   that expected results based on easily visually observed individuals or detectable with traditional

428   sampling gear such as nets may be very different from results using eDNA. This does caution that

429   eDNA surveys may have different purposes and may not be directly comparable to existing surveys

430   (Shelton et al., 2016).

431   We acknowledge that sampling artifacts may have affected our results. For example if entire

432   multicellular individuals were captured in our samples, their DNA could be in much greater density

433   than eDNA, affecting the observed community. Our sampling bottles excluded particles larger than

434   500 micrometers, but gametes and very small larvae could have gained entry. It is possible that

435   even a single small individual, containing many thousand mitochondria, would overwhelm the signal

436   of another species from which hundreds of cells had been sloughed from many, larger individuals.

437   Data on larval size distribution at the time of sampling from each species in our data set would

438   allow us to estimate the frequency of such events. Nevertheless, it is precisely the sensitivity to

439 small particles that makes the eDNA approach powerful, so we are reluctant to recommend that

440 aquatic eDNA sampling use finer pre-filtering. Instead, we emphasize the importance of designing

441 and selecting primer sets that selectively amplify target organisms. In the case of the present study,

442 in order to recover patterns matching our expectations, this would be non-transient, benthic marine

443 organisms lacking any pelagic life stage.

444   The marker we chose for this study detects a wide variety of metazoans while excluding other

445 more common taxa; however, it does not effectively discriminate among species within a higher group

446 in all cases. Other markers, such as mitochondrial cytochrome c oxidase subunit 1 (COX1, CO1,

447 or COI) may provide adequate species-level resolution in some metazoan groups, but have other

448 shortcomings including taxon dropout (Deagle et al., 2014) and amplification of more abundant non-

449 metazoans, as we discovered in an accompanying study (Kelly et al., 2016a). Both have undesirable

450 effects of biasing estimates of diversity. In our case, it is possible that the lumping of multiple

451 species into one group underestimates the true richness of the group and of the entire sample, in

452 turn obscuring true underlying patterns of diversity. In the case of COX1, well-documented primer

453 biases cause failure to amplify some taxa, particularly in mixed samples, with the same result

454 (Deagle et al., 2014). In fact, even surveys relying on traditional capture techniques (e.g. seine

455 nets) and morphological characteristics are subject to biases imposed by the sampling gear (e.g.

456 mesh size), the observer (e.g. taxonomic expertise), and organisms (e.g. morphologically cryptic

457 species). Similarly, no single molecular marker adequately and effectively samples all taxa without

458 bias (Drummond et al., 2015), and thus the choice of marker is an important and context-dependent

459 one. Until whole-genome sequencing of individual cells is a reality, the tradeoffs between taxonomic

460 breadth and resolution will continue to be problematic for metabarcoding studies, just as they are

461 for more traditional ecological survey methods (Kelly et al., 2016a).

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

463 increase the speed and throughput of DNA sequencing and sequence processing, making sense of

464 these data in a timely manner requires that natural history data be stored in standard formats in

465 centralized repositories. The rate at which we can make sense of high-throughput survey methods

466 will be limited by our ability to collate auxiliary data. Databases such as Global Biodiversity

467 Information Facility (GBIF), Encyclopedia of Life (EOL), and FishBase (Parr et al., 2014; Froese

468 and Pauly, 2016) contain records of taxonomy, occurrence, and other rudimentary data types, but

469 there is no centralized, standardized repository for even basic natural history data such as body
470 size. As NCBI's nucleotide and protein sequence database (GenBank) has facilitated transformative
471 studies in diverse fields, an ecological analog would be a boon for biodiversity science.

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

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

Acknowledgements

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

Data Availablity

Sequence Data

492 All sequence data are available from NCBI under BioProject PRJNA338801.
493 Scripts to process raw sequence data into the contingency tables used for ecological analyses can be
494 found at:

495 `https://github.com/jimmyodonnell/banzai`

496

Project Repository

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

499 `https://github.com/jimmyodonnell/Carkeek_eDNA_grid`

500 `https://doi.org/10.5281/zenodo.242976`

Sequencing Metadata

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

Life History Data

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

Analysis Scripts

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

References

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

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

516 Barnes, M. A., Turner, C. R., Jerde, C. L., Renshaw, M. A., Chadderton, W. L., and Lodge, D. M.

517   (2014). Environmental conditions influence eDNA persistence in aquatic systems. *Environmental*
518   *Science and Technology*, 48(3):1819–1827.

519   Bell, T. (2010). Experimental tests of the bacterial distance–decay relationship. *The ISME Journal*,
520   4(11):1357–1365.

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

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

524   Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L.
525   (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10:421.

526   Chamberlain, S. A. and Szöcs, E. (2013).   taxize: taxonomic search and retrieval in R.
527   *F1000Research*, 2(0):191.

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

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

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

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

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

540   Cowart, D. a., Pinheiro, M., Mouchel, O., Maguer, M., Grall, J., Miné, J., and Arnaud-Haond, S.
541   (2015). Metabarcoding Is Powerful yet Still Blind: A Comparative Analysis of Morphological and
542   Molecular Surveys of Seagrass Communities. *Plos One*, 10(2):e0117562.

543 de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le

544   Bescot, N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J.-M., Bittner,

545   L., Chaffron, S., Dunthorn, M., Engelen, S., Flegontova, O., Guidi, L., Horák, A., Jaillon, O.,

546   Lima-Mendez, G., Lukeš, J., Malviya, S., Morard, R., Mulot, M., Scalco, E., Siano, R., Vincent,

547   F., Zingone, A., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Acinas, S. G., Bork, P.,

548   Bowler, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S.,

549   Raes, J., Sieracki, M. E., Speich, S., Stemmann, L., Sunagawa, S., Weissenbach, J., Wincker, P.,

550   and Karsenti, E. (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science*, 348(6237).

551 Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., Taberlet, P., Deagle, B. E., Jarman, S. N.,

552   and Coissac, E. (2014). DNA metabarcoding and the cytochrome c oxidase subunit I marker :

553   not a perfect match DNA metabarcoding and the cytochrome c oxidase subunit I marker : not a

554   perfect match. (September).

555 Deiner, K. and Altermatt, F. (2014). Transport distance of invertebrate environmental DNA in a

556   natural river. *PLoS ONE*, 9(2).

557 Dethier, M. N. (2010). Overview of the ecology of Puget Sound beaches. In Shipman, H., Dethier,

558   M. N., Gelfenbaum, G., Fresh, K. L., and Dinicola, R. S., editors, *Puget Sound Shorelines and*

559   *the Impacts of Armoring—Proceedings of a State of the Science Workshop*, page 262.

560 Drummond, A. J., Newcomb, R. D., Buckley, T. R., Xie, D., Dopheide, A., Potter, B. C., Heled, J.,

561   Ross, H. A., Tooman, L., Grosser, S., Park, D., Demetras, N. J., Stevens, M. I., Russell, J. C.,

562   Anderson, S. H., Carter, A., and Nelson, N. (2015). Evaluating a multigene environmental DNA

563   approach for biodiversity assessment. *GigaScience*, 4(1):46.

564 Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*,

565   26(19):2460–2461.

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

567   Pfrender, M. E., Lamberti, G. A., and Lodge, D. M. (2016). Quantification of mesocosm fish and

568   amphibian species diversity via environmental DNA metabarcoding. *Molecular Ecology Resources*,

569   16(1):29–41.

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

572 Fonseca, V. G., Carvalho, G. R., Nichols, B., Quince, C., Johnson, H. F., Neill, S. P., Lambshead,
573   J. D., Thomas, W. K., Power, D. M., and Creer, S. (2014). Metagenetic analysis of patterns of
574   distribution and diversity of marine meiobenthic eukaryotes. *Global Ecology and Biogeography*,
575   23(11):1293–1302.

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

577 Guardiola, M., Uriz, M. J., Taberlet, P., Coissac, E., Wangensteen, O. S., and Turon, X.
578   (2015). Deep-Sea, Deep-Sequencing: Metabarcoding Extracellular DNA from Sediments of Ma-
579   rine Canyons. *PLOS ONE*, 10(10):e0139633.

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

583 Guo, H., Chamberlain, S. A., Elhaik, E., Jalli, I., Lynes, A. R., Marczak, L., Sabath, N., Vargas,
584   A., Więski, K., Zelig, E. M., and Pennings, S. C. (2015). Geographic variation in plant commu-
585   nity structure of salt marshes: Species, functional and phylogenetic perspectives. *PLoS ONE*,
586   10(5):e0127781.

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

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

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

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

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

597 Kelly, R. P., Closek, C. J., O'Donnell, J. L., Kralj, J. E., Shelton, A. O., and Samhouri, J. F.
598   (2016a). Genetic and manual survey methods yield different and complementary views of an
599   ecosystem. *Frontiers in Marine Science*, 3:283.

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

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

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

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

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

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

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

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

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

620 Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads.

621   *EMBnet.journal*, 17(1):10.

622 Martiny, J. B. H., Eisen, J. A., Penn, K., Allison, S. D., and Horner-Devine, M. C. (2011). Drivers of

623   bacterial beta diversity depend on spatial scale. *Proceedings of the National Academy of Sciences*,

624   108(19):7850–7854.

625 Nekola, J. C. and White, P. S. (1999). The distance decay of similarity in biogeography and ecology.

626   *Journal of Biogeography*, 26(4):867–878.

627 O'Donnell, J. L., Kelly, R. P., Lowell, N. C., and Port, J. A. (2016). Indexed PCR Primers Induce

628   Template-Specific Bias in Large-Scale DNA Sequencing Studies. *PLOS ONE*, 11(3):e0148698.

629 Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R.,

630   O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., and Wagner, H.

631   (2016). vegan: Community Ecology Package.

632 Parr, C. S., Wilson, N., Leary, P., Schulz, K. S., Lans, K., Walley, L., Hammock, J. A., Goddard,

633   A., Rice, J., Studer, M., Holmes, J. T. G., and Corrigan, R. J. (2014). The Encyclopedia of

634   Life v2: Providing Global Access to Knowledge About Life on Earth. *Biodiversity Data Journal*,

635   2(2):e1079.

636 Phillips, N. E. and Shima, J. S. (2010). Reproduction of the vermetid gastropod dendropoma

637   maximum (Sowerby, 1825) in Moorea, French Polynesia. *Journal of Molluscan Studies*, 76(2):133–

638   137.

639 Port, J. A., O'Donnell, J. L., Romero-Maraccini, O. C., Leary, P. R., Litvin, S. Y., Nickols, K. J.,

640   Yamahara, K. M., and Kelly, R. P. (2016). Assessing vertebrate biodiversity in a kelp forest

641   ecosystem using environmental DNA. *Molecular Ecology*, 25(2):527–541.

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

643 Ray, G. C. (1988). Ecological diversity in coastal zones and oceans. In Wilson, E. O. and Peter,

644   F. M., editors, *Biodiversity*, chapter 4. National Academies Press (US), Washington, DC.

645 Renshaw, M. A., Olds, B. P., Jerde, C. L., McVeigh, M. M., and Lodge, D. M. (2015). The

646   room temperature preservation of filtered environmental DNA samples and assimilation into a

647   phenol–chloroform–isoamyl alcohol DNA extraction. *Molecular Ecology Resources*, 15(1):168–176.

648 Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open

649   source tool for metagenomics. *PeerJ*, 4:e2584.

650 Sassoubre, L. M., Yamahara, K. M., Gardner, L. D., Block, B. A., and Boehm, A. B. (2016).

651   Quantification of Environmental DNA (eDNA) Shedding and Decay Rates for Three Marine

652   Fish. *Environmental Science & Technology*, 50(19):10456–10464.

653 Schnell, I. B., Bohmann, K., and Gilbert, M. T. P. (2015). Tag jumps illuminated - reducing

654   sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*,

655   pages n/a–n/a.

656 Shelton, A. O., O'Donnell, J. L., Samhouri, J. F., Lowell, N., Williams, G. D., and Kelly, R. P.

657   (2016). A framework for inferring biological communities from environmental DNA. *Ecological

658   Applications*, 26(6):1645–1659.

659 Shogren, A. J., Tank, J. L., Andruszkiewicz, E. A., Olds, B., Jerde, C., and Bolster, D. (2016).

660   Modelling the transport of environmental DNA through a porous substrate using continuous

661   flow-through column experiments. *Journal of The Royal Society Interface*, 13(119):423–425.

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

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

664 Strathmann, M. F. (1987). *Reproduction and Development of Marine Invertebrates of the Northern

665   Pacific Coast: Data and Methods for the Study of Eggs, Embryos, and Larvae.* University of

666   Washington Press, Seattle.

667 Strathmann, M. F. and Strathmann, R. R. (2006). A Vermetid Gastropod with Complex Intracap-

668   sular Cannibalism of Nurse Eggs and Sibling Larvae and a High Potential for Invasion. *Pacific

669   Science*, 60(1):97–108.

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

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

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

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

679   Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M.,
680   Solovyev, V. V., Rubin, E. M., Rokhsar, D. S., and Banfield, J. F. (2004). Community structure
681   and metabolism through reconstruction of microbial genomes from the environment. *Nature*,
682   428(6978):37–43.

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

688   Wetzel, C. E., de Bicudo, D. C., Ector, L., Lobo, E. A., Soininen, J., Landeiro, V. L., and Bini,
689   L. M. (2012). Distance Decay of Similarity in Neotropical Diatom Communities. *PLoS ONE*,
690   7(9):e45071.

691   Yang, Z. and Khangaonkar, T. (2010). Multi-scale modeling of Puget Sound using an unstructured-
692   grid coastal ocean model: From tide flats to estuaries and coastal waters. *Ocean Dynamics*,
693   60(6):1621–1637.

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

696 **Figures**
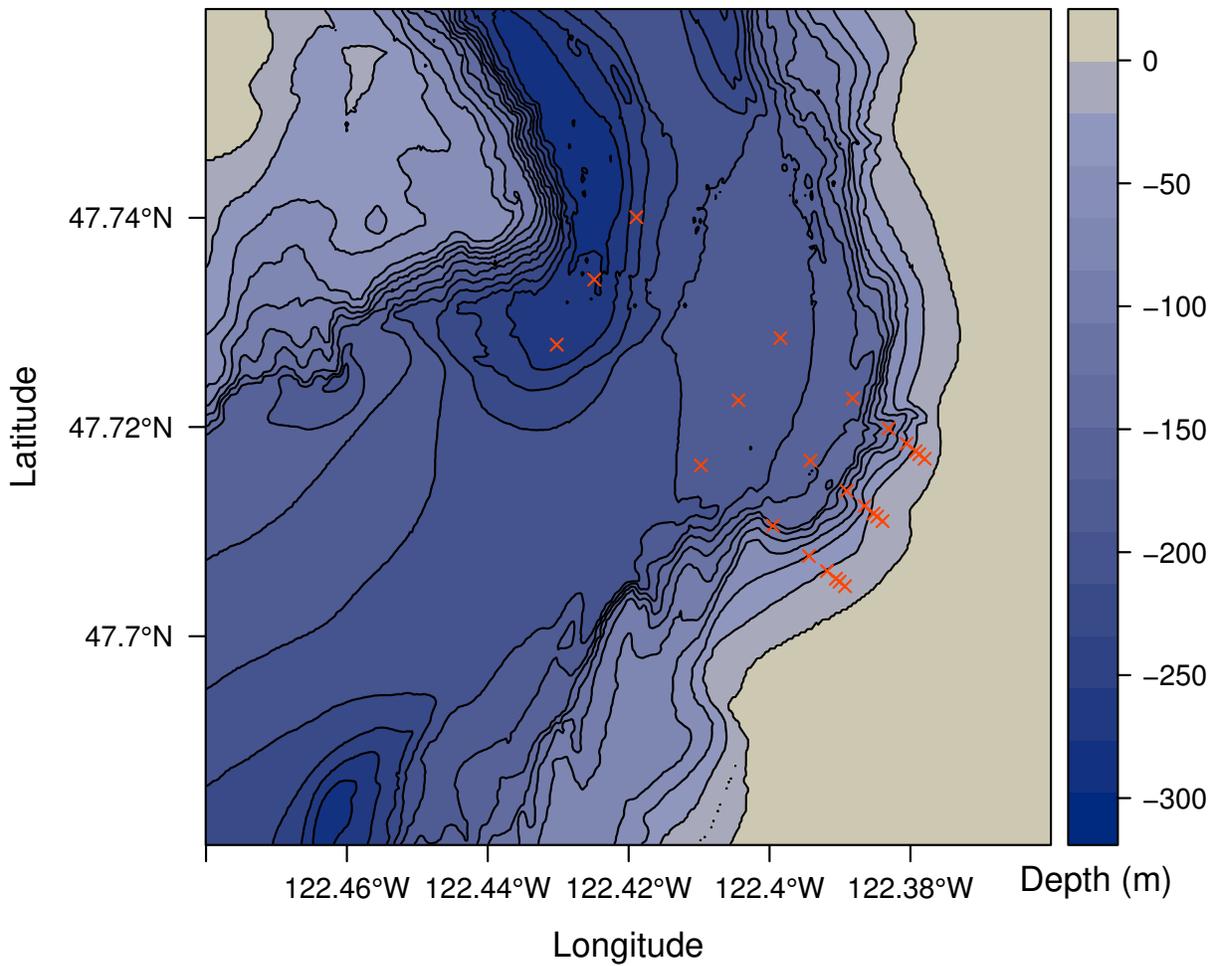
Figure 1: Map of study area. Depth in meters below sea level is indicated by shading and 25 meter contours. Sampled locations are indicated by red points.
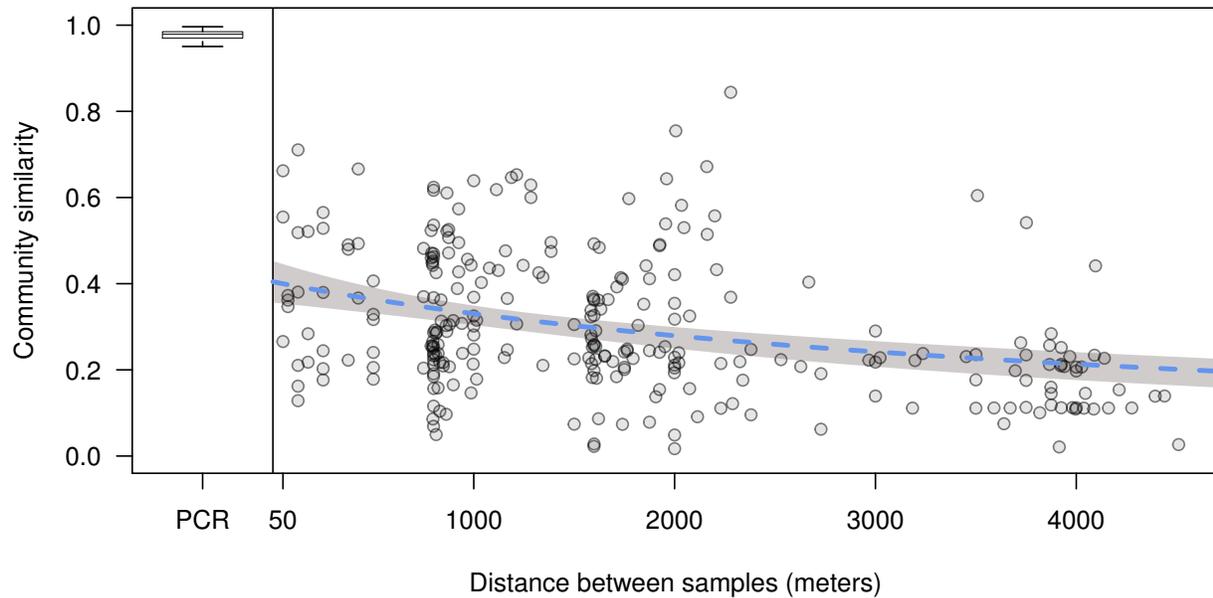
Figure 2: Distance decay relationship of environmental DNA communities. Each point represents the Bray-Curtis similarity of a site sampled along three parallel transects comprising a 3000 by 4000 meter grid. Blue dashed line represents fit of a nonlinear least squares regression (see Methods), and shading 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.
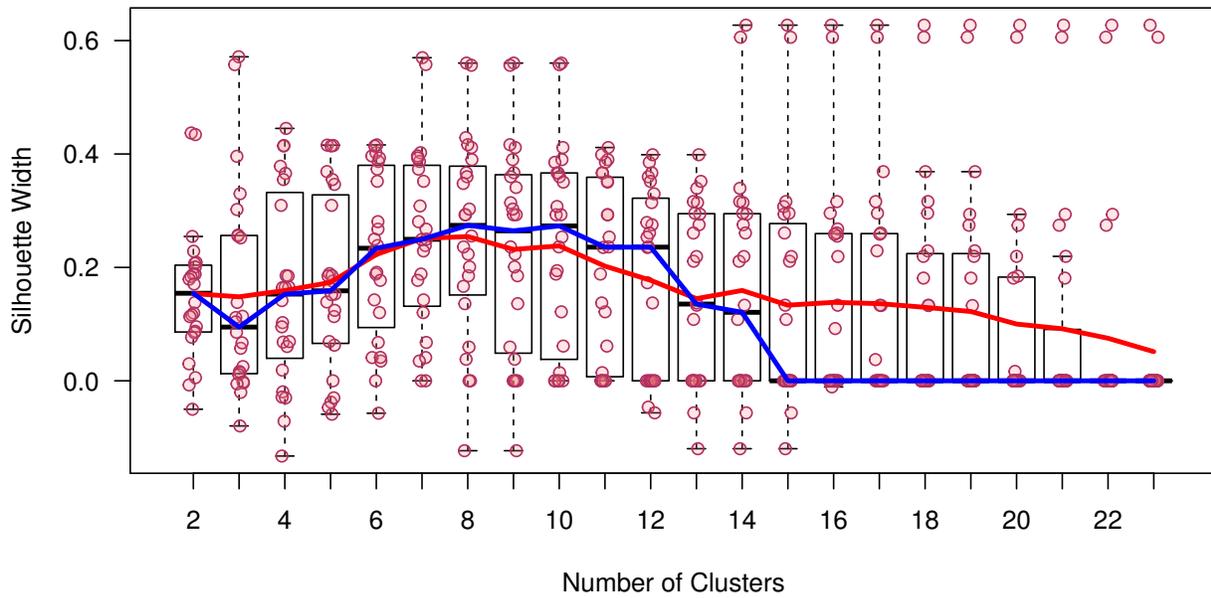
Figure 3: Silhouette widths from PAM analysis. Points are the width of the PAM silhouette of each sample at each number of clusters (K). Red line is the mean, blue line is the median. Boxes encompass the interquartile range with a line at the median, and the whiskers extend to 1.5 times the interquartile range. Boxplot outliers are omitted for clarity.

Figure 4: 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 (A; 2) and mean silhouette width (B; 8).
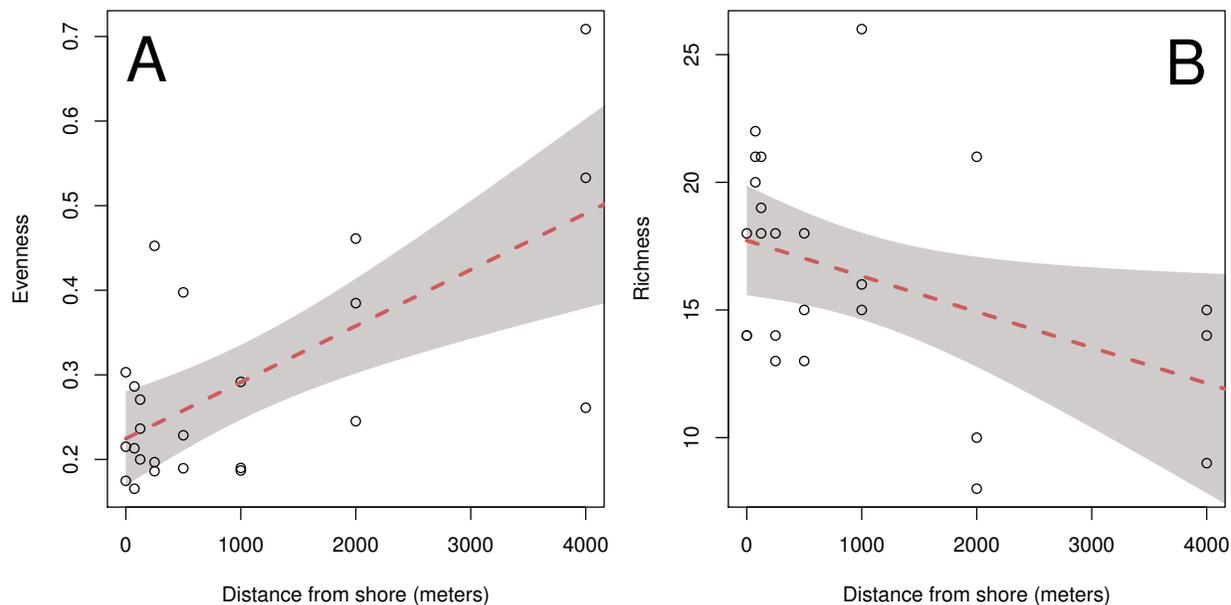
Figure 5: Aggregate diversity metrics of each site plotted against distance from shore. Both Simpson's Index of evenness (A) and richness (B) are shown, and have been computed from the mean abundance of unique DNA sequences found across 4 PCR replicates at each of 24 sites. Lines and bands illustrate the fit and 95% confidence interval of a linear model.
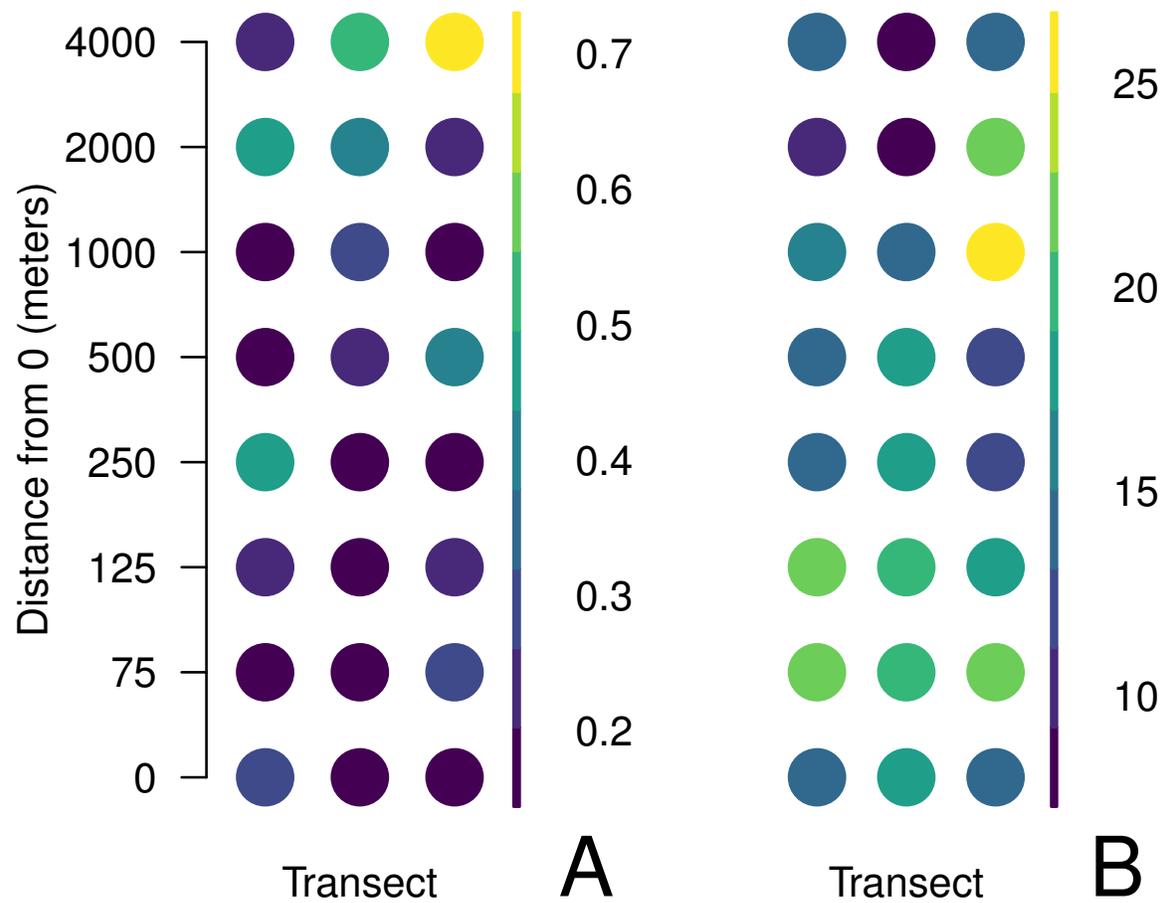
Figure 6: Aggregate measures of diversity at each sample site. Data are rarefied counts of mitochondrial 16S sequences collected from 3 parallel transects in Puget Sound, Washington, USA. Evenness (A) is the probability that two sequences drawn at random are different; richness (B) represents the total number of unique sequences from that location.
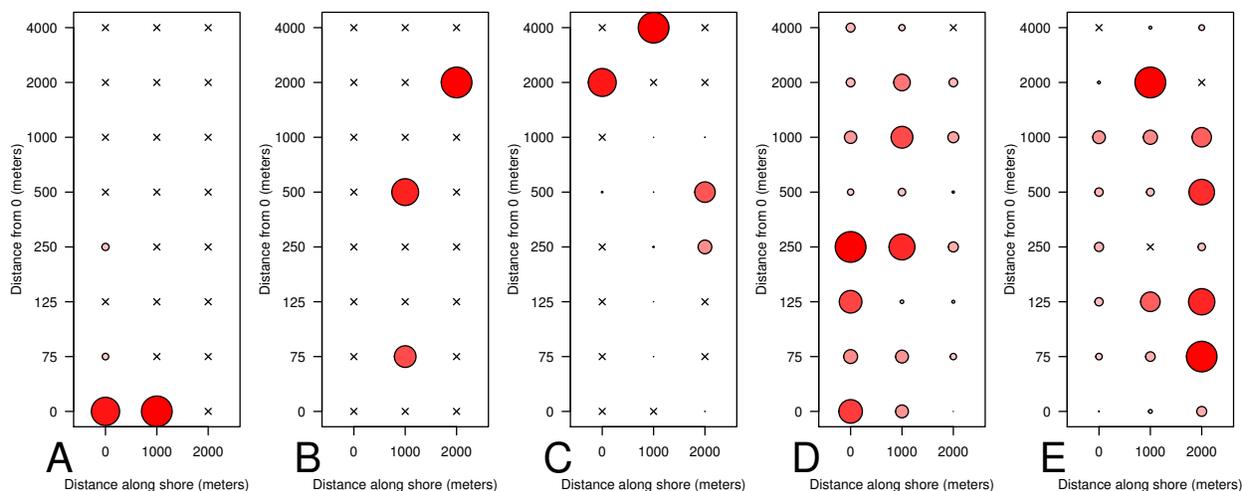
Figure 7: Distribution of eDNA from select taxa. Taxa represented are: Embiotocidae (A), Cupo-laconcha meroclista (B), Thysanoessa raschii (C), Mytilus (D), Sessilia (E). Circles are colored and scaled by the proportion of that taxon's maximum proportional abundance. That is, the largest circle is the same size in each of the panels, and occurs where that taxon contributed the greatest proportional abundance of reads to that sample.