

1 **Running head: Beach eDNA complements human observation**

2
3 **Title:**

4 **Beach environmental DNA fills gaps in photographic biomonitoring to track spatiotemporal**
5 **community turnover across 82 phyla**

6
7
8 **Authors:** Rachel S. Meyer^{1*}, Teia M. Schweizer^{1†*}, Wai-Yin Kwan¹, Emily Curd¹, Adam Wall², Dean
9 Pentcheff^{2,3}, Regina Wetzer², Eric Beraut¹, Alison Young³, Rebecca Johnson³, Robert K. Wayne¹

10 * Co-first author

- 11
12 1. UCLA Dept of Ecology and Evolutionary Biology, Los Angeles CA 90095
13 2. Natural History Museum of Los Angeles Marine Biodiversity Center, Los Angeles CA 90007
14 3. California Academy of Sciences, San Francisco CA 94118

15 † Present address: Colorado State Univ Dept of Biology, Fort Collins CO 80523

16 RS Meyer and TM Schweizer should be considered joint first author

17 * Corresponding author: 001-206-351-7997; rsmeyer@g.ucla.edu

18
19
20 **Keywords: tidepools, iNaturalist, biodiversity, holobiome, DNA metabarcoding, marine protected**
21 **areas**

22
23
24
25 **Acknowledgements:** Support for this project was provided by the University of California Catalyst
26 Program CA-16-376437. We thank the volunteers who collected CALeDNA samples and iNaturalist
27 observations. We thank M. Delaney of NOAA and thank the California Dept. of Fish and Wildlife for
28 assistance with permits and permissions. We thank M. Lin, and Z. Gold for assistance with analyses. The
29 authors (AW, NDP, RW) graciously acknowledge NHM Marine Biodiversity Center infrastructure support
30 from colleagues, Kathy Omura and Jenessa Wall. This is Contribution Number 5 of the NHM Diversity
31 Initiative of the Southern California Ocean.

32
33 **Author Contributions:** RSM, TMS, EC, AY, RJ, and RKW designed the study, with WK and AW designing
34 and performing additional analyses. RSM, TMS, EC, and RKW generated the eDNA data, DP and RW
35 contributed DNA reference data and analyzed taxonomic results. RSM, EB and WK managed the data. All
36 authors contributed to analyses and to writing the manuscript.

44 **Abstract:**

45 Environmental DNA (eDNA) metabarcoding is emerging as a biomonitoring tool available to the citizen
46 science community that promises to augment or replace photographic observation. However, eDNA
47 results and photographic observations have rarely been compared to document their individual or
48 combined power. Here, we use eDNA multilocus metabarcoding, a method deployed by the CALeDNA
49 Program, to inventory and evaluate biodiversity variation along the Pillar Point headland near Half Moon
50 Bay, California. We describe variation in presence of 13,000 taxa spanning 82 phyla, analyze
51 spatiotemporal patterns of beta diversity, and identify metacommunities. Inventory and measures of
52 turnover across space and time from eDNA analysis are compared to the same measures from Global
53 Biodiversity Information Facility (GBIF) data, which contain information largely contributed by iNaturalist
54 photographic observations. We find eDNA depicts local signals with high seasonal turnover, especially in
55 prokaryotes. We find a diverse community dense with pathogens and parasites in the embayment, and a
56 State Marine Conservation Area (SMCA) with lower species richness than the rest of the beach
57 peninsula, but with beta diversity signals showing resemblance to adjacent unprotected tidepools. The
58 SMCA differs in observation density, with higher density of protozoans, and animals in Ascidiacea,
59 Echinoidea, and Polycladida. Local contributions to beta diversity are elevated in a section of East-facing
60 beach. GBIF observations are mostly from outside the SMCA, limiting some spatial comparisons.
61 However, our findings suggest eDNA samples can link the SMCA sites to sites with better GBIF inventory,
62 which may be useful for imputing species from one site given observations from another. Results
63 additionally support >3800 largely novel biological interactions. This research, and accompanying
64 interactive website support eDNA as a gap-filling tool to measure biodiversity that is available to
65 community and citizen scientists.

66

67 **1. Introduction:**

68
69 The 6th mass extinction now underway is projected to affect a million species (Diaz et al., 2019), and
70 consequently, scientists and the public need to catalyze the development of next-generation
71 biodiversity monitoring strategies for a comprehensive inventory of biodiversity (National Research
72 Council, 2001). Our interest is in a strategy that activates a public constituency with access to the latest
73 digital and molecular tools and an understanding of the data and results so they can innovate solutions
74 together with the scientific community. Volunteers for citizen and community science outnumber
75 professional scientists 18 to 1 (Groom et al., 2017), and have contributed a substantial amount of
76 observational data points that are now used to track biodiversity variation over space and time (Bird et
77 al., 2014). One major source of observations is the iNaturalist community science platform [Cal Academy
78 of Sciences (CAS) and National Geographic Society], which has exceeded 12 million ‘research grade’, or
79 photographed and validated, records. For certain species and localities, these observations comprise the
80 bulk of data points over other occurrence data, such as museum and voucher specimens. iNaturalist and
81 other volunteer-collected species occurrence records (such as eBird) help populate the Global
82 Biodiversity Information Facility (GBIF; Robertson et al., 2014), which also includes collection and
83 observation data curated by thousands of institutions (n=1332 ‘publishers’ as of January 11, 2019).
84 Citizen and community scientists are eager to be data contributors and deepen their involvement in
85 knowledge assessment and possible conservation actions, (Hecker et al., 2018). Everyone is impacted by
86 the challenges of biodiversity loss.

87
88 Environmental DNA (eDNA) presents an attractive method for obtaining alternative biodiversity data.
89 The method is considered non-invasive as small amounts of soil, sediment, or water are sufficient for
90 DNA extraction, which can be easily collected by volunteers or passively collected with unmanned
91 devices. eDNA can be subjected to multilocus metabarcoding and next-generation sequencing,

92 producing snapshots that span all kingdoms of life. However, because we don't have sufficient DNA
93 barcodes for all species, sometimes there are insufficient data to resolve taxonomy or avoid
94 mismatches. DNA research in the environment is also limited by the non-even distribution of DNA from
95 taxa (especially inshore and terrestrial environments; O'Donnell et al. 2017), where small bodied and
96 more evenly distributed species have better detectability over large mobile organisms (Zinger et al.
97 2019a). Therefore, eDNA-based biodiversity inventories may benefit from complementation with other
98 evidence-based observations such as photos or specimens. Nonetheless, the sheer data-richness of
99 eDNA may elucidate robust metacommunities with dynamics that inform ecosystem management (e.g.
100 Peters et al., 2019) and that may even allow imputation about the presence of larger organisms from
101 their associated smaller taxa comprising their holobiomes (i.e. the total genomes in and on a eukaryotic
102 organism, possibly functioning together as evolutionary units; Guerrero et al., 2013). However, as a
103 nascent field, this topic has hardly been explored.

104
105 Photographic observations present challenges that systematic collection of eDNA may help resolve.
106 Diversity estimates are limited to where people choose to go, to what people decide to photograph or
107 sample, or to where camera traps are installed. Further, by the nature of photographs, it is mainly
108 relatively large or morphologically apparent, diurnal organisms. How can eDNA surveys augment or
109 predict taxa in observational databases? If eDNA provides a next-generation biodiversity assessment
110 tool, how can it be integrated into biomonitoring by the public and scientists? By using eDNA samples to
111 fill information gaps and stimulate broader human observation, can we mitigate taxon (e.g. plant
112 blindness; Allen, 2003) and visitation bias?

113
114 This study presents an analysis of eDNA metabarcoding collected through community science effort of
115 the CALeDNA program (www.ucedna.com; Meyer et al., 2019). Our analysis focuses on an inventory of

116 biotic communities on the Pillar Point headland California beach that has been previously been
117 intensively surveyed by the California Academy of Sciences (CAS) Citizen Science team
118 (<https://www.inaturalist.org/projects/intertidal-biodiversity-survey-at-pillar-point>). The beach has a
119 marshland and harbor graded C-F for water quality on the Heal the Bay Beach Report Card (2016;
120 healthebay.org), signaling it is polluted and potentially under threat. It also has a State Marine
121 Conservation Area (SMCA) with restricted fishing, that is difficult to access by land, which received
122 grades A-C. It remains unclear how biologically different these areas are. We compare eDNA results to
123 human observation records (GBIF; largely including iNaturalist research grade observations) to
124 determine their complementarity and the value of each to fill gaps where data are limited across space
125 and time. We find eDNA and photographs have little overlap, with eDNA covering lower trophic.
126 Systematic collection of eDNA can describe differences spatial and temporal biodiversity variation of the
127 beach, and provides a rich view of the community ecological network, including holobiomes. Even
128 though this is only a single case study, our results suggest that eDNA may be useful to find surrogate
129 sites where GBIF data is readily collected, which may be useful proxies for diversity in sites that are
130 difficult for the volunteer community to monitor with photographs. With community empowerment as
131 the goal, results in this paper are paired with interactive web pages for the public, who are potential
132 eDNA adopters, to explore (https://data.ucedna.com/research_projects/pillar-point?).

133

134 **2. Materials and Methods:**

135 Detailed methods including commands for running programs are in the Supplemental Methods.

136

137 *2.1. Sample collection for eDNA*

138 Samples of soil, sand and sediment for eDNA analysis were collected at Pillar Point by a total of fifteen
139 community scientists and three UCLA researchers during low-tide windows of three days: February 8,

140 2017, April 29, 2017 and April 30, 2017. Our definition of community scientist is a volunteer participating
141 in a science project, which could be designed and empowered by non-professional scientists or
142 organized by professional scientists [see Meyer and Drill (2019) for further discussion]. The February
143 sampling date coincided with the monthly Pillar Point survey organized by CAS Citizen Science Academy
144 staff, which selects seasonally low tides for observation. In February, most eDNA samples were only
145 collected near or above the mean high water line (aka mean high tide mark; MHT) because of permit
146 restrictions. Most sites below the MHT line were sampled in April, as were all SMCA samples. Collections
147 were made under permit MBNMS-2017-019 issued by the NOAA Office of National Marine Sanctuaries.
148 Sample metadata are provided in Table S1.

149
150 For the eDNA collection, sample collectors were organized into groups to cover different areas of the
151 Pillar Point headland. Sample collectors then spread out and selected sites that looked like a typical
152 representation of the local environment. Surface soil, sand, or submerged sediment were collected
153 following the CALeDNA community science program instructions, which required fresh gloves to be
154 worn at each sampling site, three 'biological replicates' of 2 mL cryotubes to be filled with substrate
155 approximately 30 cm apart from each other. Sampling was limited to the exposed (top 3 cm layer)
156 surface or submerged sediment no deeper than could be reached at arm's length. Collectors used a
157 Kobo Toolbox phone webform (<https://ee.kobotoolbox.org/x/#Y1sl>) to record sampling time,
158 geolocation, photographs of the sampling site, and other metadata. Metadata for each sample are
159 available through data.ucedna.com. Photographs were later used to confirm environmental metadata
160 (Table S1) for the sites. All collection instructions, data and photographs are published online in the
161 CALeDNA database at www.ucedna.com.

162

163 *2.2. Assembling observation records*

164 GBIF records from all contributors were downloaded September 4, 2018 from GBIF using drawn
165 polygons on a map of Pillar Point (coordinates and download DOI in Appendix 1). eBird observations
166 were removed because they were concentrated only in the embankment. A total of 13,924 occurrence
167 records were retained. The polygons correspond to the headland embayment, unprotected reef area,
168 and SMCA protected area (data in Table S2).

169

170 *2.3. eDNA metabarcoding library preparation and sequencing*

171 Subsamples of equal mass of all three biological replicates were pooled and homogenized. A ~0.25g
172 amount of the subsample was extracted with the Qiagen DNeasy PowerSoil Kit according to
173 manufacturer's instructions (Qiagen, Valencia, CA, USA). Kit details provided by the manufacturer
174 demonstrate retrieval of amplifiable DNA from bacteria, fungi, algae, and animals. A blank PowerBead
175 tube was extracted alongside each batch of ~10 samples as a negative extraction control. Eight DNA
176 extraction controls were pooled into a single sample used in all PCR amplifications.

177

178 Sample DNA was amplified for targeted taxa using four primer sets based on the following universal
179 primers with a Nextera adaptor modification: MiFish *12S* (Universal Teleost; Miya et al. 2015), *16S* (515D
180 and 806R; Caporaso et al., 2012), *18S* (V8-V9; Amaral-Zettler et al., 2009; Bradley et al., 2016), and the
181 partial sequence of *Cytochrome Oxidase Subunit 1 (CO1)*; mlCOLintF and jgHCO2198; Leray et al., 2013).
182 Primer sequences are included in Supplemental Methods. Three PCR replicates were run per amplicon
183 per sample, and these were pooled and converted to Nextera libraries (details in Supplemental
184 Methods), and sequenced using a MiSeq Illumina Next Generation Sequencer with Reagent Kit V3 (2 x
185 300 bp) at a goal depth of 25,000 reads in each direction per marker per sample. The original run
186 produced reverse reads with lower than standard quality, and so the run was repeated. Both runs were
187 analyzed separately and resulting tables were merged.

188

189 *2.4. eDNA classification*

190 Demultiplexed library Fastq files were processed with the *Anacapa Toolkit* (archived version
191 (doi:10.5281/zenodo.3064152)(github.com/limey-bean/Anacapa; Curd et al., 2019) using default
192 parameters. In brief, reads were processed first by *cutadapt* (Martin, 2011) to remove adapters and any
193 3' primer reverse complement. Reads were quality trimmed with the *FastX-Toolkit*
194 (http://hannonlab.cshl.edu/fastx_toolkit/) and only reads with a minimum length over 100 bp were
195 retained. *Cutadapt* was used to sort reads by primer and then used to remove primers. Read bins were
196 denoised, dereplicated and merged or retained as paired unmerged, forward-only or reverse-only reads
197 using *dada2* (Callahan et al., 2016). Chimeras were additionally removed using *dada2*. This produced
198 amplicon sequence variants (ASVs) that were then processed using *Bowtie 2* (Langmead and Salzberg,
199 2012) to query custom reference databases (described in the following paragraph) and determine up to
200 100 reference matches with a minimum percent coverage of the sample read set to 70% for *CO1*, 80%
201 for all others and a minimum percent identity of 70% for *CO1*, and 90% for all others. Details on how
202 these settings were chosen are in Supplemental Methods and Table S3. Reads per taxon were compiled
203 for each marker using a minimum bootstrap classifier confidence (BCC) = 60 for *16S*, *12S*, and *18S*, and a
204 minimum BCC = 70 for *CO1*.

205

206 Reference databases for *16S*, *18S*, *CO1*, and *12S* were generated with the *CRUX* step of the *Anacapa*
207 *Toolkit* (Curd et al., 2019 ;DOI: <https://doi.org/10.5061/dryad.mf0126f>) that queried NCBI nr/nt
208 databases. The *CO1* database was subsequently modified by appending all BOLD *CO1-5P* sequences,
209 accessed via the BOLD API on September 24, 2018 and comprising 5.08 million sequences, and also
210 appending 847 recently generated invertebrate *CO1-5P* barcode sequences from the LACM DISCO
211 program (subsequently published in BOLD with the same identifiers). The *CO1-5P* region is the most

212 commonly used CO1 locus, described first by Folmer et al (1994) and it includes our amplified region.

213 These reference databases are permanently archived [*in DRYAD – link pending*].

214

215 BOLD and NCBI use different classifications systems for higher taxonomy (order to superkingdom).

216 Results tables originally had taxon strings that could be a mixture of classifications. We united them

217 using the higher classification by Ruggiero et al. (2015; online version) for all results tables and merged

218 rows with identical names by summing column read counts. This procedure was only used for

219 decontaminated results.

220

221 For *16S*, *18S*, and *12S* markers, we statistically removed reads and taxa considered contaminants using

222 the program *Decontam* version 1.1.2 (Davis et al., 2018) with prevalence method and the threshold

223 setting of 0.1. Taxa with only a single total read were subsequently removed in the final results (Tables

224 S4-S7).

225

226 For *CO1*, because there were few taxa that overlapped between real samples and extraction or PCR

227 negative controls, results tables were decontaminated simply by removing any taxon that was present in

228 one or more reads in controls. The extraction and PCR negative controls contained a total of three taxa,

229 two being single reads in either control or real sample, and one that did occur as 9 reads in a control and

230 in more reads in four real samples, which was *Nutricula tantilla* Gould, a very small clam. We did not

231 include this taxon in results, and note it was not in GBIF Pillar Point data but is known from California

232 (GBIF.org). Singletons were also subsequently removed.

233

234 Using the *CO1* results, we further examined possible contaminants by heeding Zinger et al., (2019b) who

235 describe the various contamination issues, specifically, the risk of index hopping among samples in the

236 same MiSeq run. We looked for cross-contamination between the Pillar Point samples, which were the
237 ~54% of the identified reads on the MiSeq run, and California Vernal Pool samples which were largely
238 the remainder of the run, and found only 43 taxa overlapped, 15 of which were singletons in the Pillar
239 Point samples, and 34 of which were soil-dwelling Ascomycota or amoeba that could theoretically occur
240 in both habitats (Table S6). Not one read of known vernal pool-specific animals, which each had >2500
241 reads in the vernal pool samples, were recovered in Pillar Point samples (e.g., *Gammarus cf. fossarum*,
242 or *Alona*, which are usually found in ephemeral freshwaters and known from vernal pools; Keeley and
243 Zedler, 1998; Hupalo et al., 2018). We also counted the amount of PhiX in sample libraries using BBDuk
244 (BBMap – Bushnell B. – sourceforge.net/projects/bbmap/), and found it was (<8e-05) in each library. We
245 found these low occurrences of contamination from extraction or index-hopping suggested further
246 decontamination steps were unnecessary. The *Anacapa* output results including the non-Pillar Point
247 samples, and files detailing sequences themselves and other details of taxonomy assignment
248 confidence, are provided for *CO1* in Table S6.

249
250 Alpha diversity analyses used a stricter filter of a minimum of five reads to reduce some inflation effects
251 from ASVs. All analyses requiring rarefaction used settings retaining 2000 reads for *18S* or *CO1*, and
252 4000 reads for *16S*, aiming for a cutoff above the linear growth zone of the curve. Rarefaction curves
253 used to inform the sequence number for rarefaction are shown in Figure S1.

254
255 *2.5. Global biotic interactions (GloBI)*

256 We downloaded all of the open access GloBI species interactions database (Poelen et al., 2014;
257 <http://globalbioticinteractions.org>) on September 2, 2018, totaling 3,293,470 records. These data were
258 used in two tests. First, we retrieved all the species that had 10 or more observations from the Intertidal
259 Biodiversity Survey at Pillar Point in iNaturalist (<https://www.inaturalist.org/projects/intertidal->

260 [biodiversity-survey-at-pillar-point?tab=species](#)), and searched for all GloBI records of interactions
261 involving these frequently observed species ([Biotic Interactions](#)). We screened the list of interacting taxa
262 for overlap with GBIF species and eDNA species and reported these. Second, we summarized all GloBI
263 interactions to family level, and recorded the interaction type. We used the major categories of “eats”
264 or “interacts_with” to tally the frequency with which the families detected in eDNA or GBIF are on
265 source or sink for these interactions, and used these sums in Chi-Square tests in *R* (version 3.5.0) to test
266 for overrepresentation of interaction source or sinks in eDNA and GBIF results.

267

268 2.6. Statistical analyses for diversity and enrichment

269 The *16S*, *18S*, and *CO1* results (Tables S4-S7) were used to generate separate alpha diversity and beta
270 diversity plots, and used to calculate statistics including Local Contributions to Beta Diversity (LCBD) and
271 Analysis of Similarity (ANOSIM). To prepare results for these tests and plots, decontaminated Anacapa
272 results tables and metadata were converted to *Phyloseq* objects (McMurdie and Holmes, 2013) using
273 the `convert_anacapa_to_phyloseq` function in the *Ranacapa* version 0.1.0 (Kandlikar et al., 2018) *R*
274 package in *R* version 3.5.0. *Phyloseq* and *microbiomeSeq* *R* package version 0.1

275 (<https://github.com/umerijaz/microbiomeSeq.git>) were used to calculate richness, Simpson, and
276 Shannon alpha diversity and test for significant differences among groups with ANOVA set to $P=0.05$.
277 Plots were made using *ggplot2* (Wickham 2009). Full scripts are provided in Supplemental Methods.

278

279 Beta diversity as relative abundance barplots and LCBD statistics were generated in *R* using the addition
280 of the *adespatial* (multivariate multiscale spatial analysis) package version 0.3.2 function selecting the
281 Hellinger dissimilarity coefficients method (Dray et al., 2016). ANOSIM beta diversity statistics and
282 distance plots were computed using both Jaccard and Raup-Crick dissimilarity indices (Raup and Crick,
283 1979). Both treat the data as presence/absence but Raup-Crick considers underlying alpha diversity, as

284 calculated in the Vegan vegdist method (Chase et al., 2011; Oksanen et al. 2018) in *Phyloseq*. Principal
285 Coordinate Ordinations were made using the Jaccard method.

286

287 We performed density enrichment tests in *R* using the following approach. Taxon results tables were
288 summarized by class and converted to sample presence/absence tables. Kruskal-Wallis rank sum tests
289 (Hollander, et al., 2013) were performed to identify classes with significant differences among zone
290 groups. Then, significant classes were subjected to the post-hoc Dunn test (Dunn, 1964) using the
291 Benjamini-Hochberg (Benjamini and Hochberg, 1995) correction for multiple testing.

292

293 *2.7. Community ecological network analyses*

294 Network analysis was performed using the *SpiecEasi* version 1.0.2 package (Kurtz et al., 2015; Tipton et
295 al., 2018) in *R* with the *18S*, *16S*, and *CO1* results tables summarized to the highest resolution
296 classification of family. We did not include *12S* because the taxon list was short. Family-level results
297 tables were filtered to retain only taxa minimally present in 15% of the sites. Results from each marker
298 were processed separately, and then *16S* was co-analyzed with each of the dominantly eukaryotic
299 markers *18S* and *CO1*. Settings and commands are in Supplemental Methods, as are commands for
300 plotting networks using *Phyloseq* and *iGraph* (Csardi and Nepusz, 2006). Networks were additionally
301 plotted as interactive figures using Flourish Studio (<https://app.flourish.studio>; London, UK), given a
302 stable DOI, and were hosted on our Pillar Point project website

303 (https://data.ucedna.com/research_projects/pillar-point?). Interactions were also deposited in GloBI (
304 https://github.com/beraute/Pillar_Point_16S_18S, https://github.com/beraute/Pillar_Point_CO1_16S).

305

306 We asked what proportion of observed eDNA network interactions were previously published, and
307 tested this by comparing *18S* network taxa with 10 or more edges (highly networked) to interactions
308 published in the GloBI database (Section 2.5). Lists were assembled and compared in Microsoft Excel.

309

310 2.8. Reduction of holobiome effects

311 For the *18S* results, we tested the influence of DNA swamping from organisms and their associated
312 community (holobiome) on local contribution to beta diversity (LCBD) scores. We screened the *18S*
313 rarified DNA results for taxa with >50% proportion of the read abundance. Community ecological
314 networks (Section 2.7) were mined for these organisms and their linked taxa, and all these taxa were
315 filtered out to produce new results tables. These holobiome-filtered tables were used in LCBD
316 calculations as described in Section 2.6.

317

318 2.9. GBIF and eDNA comparisons

319 GBIF and eDNA results use different classification systems. The classification at each hierarchical level
320 was converted to a common NCBI-style taxonomy using the Global Names tool
321 (<http://globalnames.org>). This tool has been shown to increase the success of cross-mapping taxon
322 names to up to 90% across diverse databases (Patterson et al., 2016); in accordance with this number,
323 under 10% of taxa had to be dropped because they could not be converted. This slightly reduced set of
324 converted names is presented on the web platform.

325

326 2.10. Development of the web interface and comparative summary statistics

327 A web platform was created linking Squarespace, Amazon Web Services, and Heroku. Numerous data
328 visualization tools such as leaflet were used to create the user interface, and images were scraped from

329 Encyclopedia of Life (eol.org), GBIF (gbif.org) iNaturalist and Wikipedia. All scripts to generate the pages
330 are open source and available on Github at <https://github.com/UCcongenomics/caledna>.

331

332 **Results**

333

334 *3.0. Orientation*

335 Pillar Point headland contains a 1 km East-facing stretch that spans a small protected marsh with
336 agricultural runoff feeding into an embayment that contains a harbor with primarily fishing and
337 recreational crafts (Figure 1). The marsh supports a diversity of seabirds and shorebirds. The entire area
338 is bounded by a stone embankment (breakwater). A 0.4 km stretch of South-facing ‘outer’ beach has an
339 extended 0.3 km tidepool accessible during very low tides that receives heavy recreational traffic and
340 visits from school groups as well as commercial and recreational collection of mollusks and fish. North of
341 the tidepools on the West ‘outer’ beach (Figure 1) is a 0.8 km stretch that includes the Pillar Point State
342 Marine Conservation Area (SMCA). The SMCA places limits on recreational and commercial fishing, but
343 not access for recreational users (Marine Life Protection Act of 1999).

344

345 We examined the inventory of taxonomic biodiversity at Pillar Point beach from eDNA results and GBIF
346 records, grouping data in several ways. The beach areas were divided into three polygons drawn on a
347 map: 1) unprotected “embayment”; 2) “unprotected” exposed beach and tidal pools; and 3) “protected”
348 exposed beach of the SMCA (Figure 1A, 1B). We also divided the beach into five zones: 1) area exposed
349 to marsh runoff; 2) unprotected inner beach; 3) unprotected outer beach; 4) unprotected tidepools; and
350 5) protected outer beach (SMCA). The total taxa observed (Figure 1D) were used to compare
351 biodiversity among the polygons. We added metadata for each sampling site that included the month
352 collected, the position relative to the mean-high-tide (MHT) line, and the substrate.

353
354 Comparative analyses of eDNA (Tables S4-S7) and GBIF results (Appendix 1; Table S2) assessed species in
355 common, differences in estimates of taxon distribution, richness, beta-diversity within and among areas
356 of the beach (Figure 1C-F). We also evaluated the kinds of biotic interactions between taxa in both types
357 of inventory. As a community science project, we presented interactive results to engage non-scientists
358 and scientists alike. Those resources are available in this web platform:

359 https://data.ucedna.com/research_projects/pillar-point and are permanently archived in Dryad and
360 Zenodo [PENDING MANUSCRIPT ACCEPTANCE]. In the Results, we highlight the website features
361 intended for self-driven evaluation of eDNA and GBIF biomonitoring. The web platform explorable
362 features are described in Table 1.

363

364 *3.1. eDNA taxonomic inventory and richness*

365 Multilocus metabarcoding for four gene loci targeted Bacteria and Archaea (*16S*), animals and algae
366 (*CO1*), eukaryotes (*18S*), and fish (*12S*). *CO1* and *18S* partially overlap with regard to organisms
367 sequenced. Libraries for 88 sites and negative controls were sequenced in two runs: 3.34 Gb of fastq
368 data were generated in the first run (170818_300PE_MS1) and 12.25 Gb were generated in the second
369 run (171006_300PE_MS1)(NCBI SRA *PENDING ACCEPTANCE*). Results were output from the Anacapa
370 Toolkit (Curd et al., 2019) summarized to the Least Common Ancestor, and then taxa were removed that
371 were found in negative controls as well as taxa with only one read across all samples, yielding 1468 total
372 taxa for *CO1*, 2689 for *18S*, 2593 for *16S*, and 43 for *12S* (Tables S4-S7). The total unique taxa assigned to
373 kingdoms across all eDNA results were 1132 Animalia, 27 Archaea, 2,533 Bacteria, 1,588 Chromista, 516
374 Fungi, 433 Plantae, and 154 Protozoa, with 3 unassigned to a kingdom.

375

376 Alpha diversity, defined by Whittaker (1972), is a measure of species richness of a place. DNA
377 metabarcoding data has been routinely used to estimate alpha diversity for a decade (Fonseca et al.,
378 2010), and can be used for measuring genetic diversity as well as taxonomic diversity. We focus our
379 biodiversity analyses on taxonomic diversity, presenting alpha diversity analyses and its interpretation in
380 Appendix 2, but include an assessment of how genetic and taxonomic diversity are related (Table S8). In
381 taxonomic alpha diversity analyses, we found seasonal and spatial differences in both prokaryotes and
382 eukaryotes (Appendix 2; Figure S2).

383
384 We asked how much sampling month (February or April) changed the families detected. Across the total
385 1,388 families detected by the four markers, the average frequency a family was found was 18% (16 out
386 of the 88 sites). The Pearson correlation (ρ) between the normalized frequencies of these 1388 families
387 observed in February with April was 0.81, but in repeating the test using a simple tally of whether
388 families were observed at all in February or April, the correlation was $\rho=0.27$. These results, taken
389 together with the significant spatial and temporal differences found in alpha diversity analyses
390 (Appendix 2) support eDNA signals are spatially local and temporally restricted.

391

392 *3.2. eDNA-based beta diversity*

393 Analysis of Similarities (ANOSIM) beta diversity tests were performed to evaluate differences in the
394 groupings of polygon, month, position in relation to the MHT line, and zone. Raup-Crick dissimilarity
395 (Raup and Crick, 1979) detected fewer significant groupings than the Jaccard method (Table S9). Raup-
396 Crick dissimilarity results showed that month was significant for all markers and had a large effect size:
397 *18S* ($R^2=0.76$), *16S* ($R^2=1$), and *CO1* ($R^2=0.53$). The position relative to the MHT line was significant for
398 *16S* and *18S*, but effect size was small ($R^2<0.08$). Polygon was significant for *18S* and the effect size was
399 $R^2=0.35$.

400

401 Jaccard dissimilarity spatial ordination analyses showed month varied predominantly along the first axis
402 for *18S* and *16S* but not *CO1* (Figure 2). Jaccard analysis grouped by zone showed the SMCA protected
403 polygon, only sampled in April, clustered with other April samples from the unprotected tidepools
404 (Figure 2). We also observed that for *16S* results, inner beach samples clustered with other zones, falling
405 either with marsh runoff samples, or as a group surrounding a dense cluster of unprotected outer beach
406 samples. This suggests the inner beach is a zone that is heterogeneous in bacteria and archaea
407 communities, experiencing some influence from adjacent areas as evidenced by their clustering.

408

409 *3.3. Interpreting and correcting Local Contributions to Beta Diversity*

410 Beta diversity estimates describe the community composition and stability across localities sampled.
411 Barplots coloring the 21 taxa with highest relative abundance show structured community composition
412 by month, polygon, and placement relative to the MHT line (Figure 3; Figure S3).

413

414 High relative abundance of DNA from some taxa may produce a 'swamping effect' that shrouds signal
415 from other taxa (see discussion in Weber et al., 2017). In the *18S* barplot, 13 samples were severely
416 swamped with DNA (>50% of the bar; Figure 3, Figure S3), from one of several taxa, belonging to
417 anemone, worm, ciliate, diatom, and others. Samples exhibiting DNA swamping had higher Local
418 Contributions to Beta Diversity (LCBD) than the rest (Table S10), a measure of the uniqueness of one
419 sample compared to the rest of its group (Legendre and Caceres, 2013). LCBD has been increasingly
420 valued for conservation applications where researchers recommend protecting sites with extreme low
421 and high LCBD (da Silva et al., 2018). In the *18S* results of eukaryotes, there were 20 sites with significant
422 higher or lower LCBD compared to their neighbors, 7 of which met our criteria as swamped.

423

424 We hypothesized the stability of LCBD scores was sensitive to swamping from the most abundant taxa
425 and their associated holobiome communities, and tested this with the 18S results (Appendix 3). DNA
426 signals from the swamping taxon were removed from results along with all of their associated taxa
427 linked with that organism in downstream community ecological networks (Section 3.6), and then LCBD
428 scores were recalculated. Holobiome reduction did produce lower LCBD scores for 19 samples, and only
429 1 of the 13 originally swamped samples remained significantly different from the rest of its group (Table
430 S10; Figure S4). The total number of significantly different samples was reduced from 20 to 9, and all
431 had higher LCBD than others in their group.

432
433 Eight of these nine high LCBD samples were from the cluster of the inner beach, south of the marsh. This
434 result suggests this East-facing beach, a rarity in California, may harbor distinct communities. This inner
435 beach is more protected from storms than outer beaches, though may be protected from marsh runoff
436 by the breakwater. We did additional LCBD analyses to explore the similarity of values across markers
437 (Appendix 3).

438
439 *3.4. Enrichment tests reveal density patterns that characterize each zone*

440 To understand biological variation at the level of spatial density, which is relevant to natural areas
441 management, we examined which classes of taxa across kingdoms differed in their density of detection
442 among the Pillar Point beach zones (Figure 4; Table S11). We tallied the samples within zones with DNA
443 signal from a class, and used Kruskal-Wallis and Dunn post-hoc tests to test for enrichment in the
444 density a class was detected. Because these results have management implications, we scrutinized our
445 taxonomic assignments to class using phylogenetic analysis and BLAST queries to vet our trust in the
446 correct class assignment (data not shown). We found two classes that could not be clearly delimited
447 from other classes in their respective phyla: the class Crinoidea, which typically occurs in deeper marine

448 environments could not be delimited from other Echinoderms. This may be due to size variation in *CO1*
449 sequences for this clade, and sequence variation shared with other organisms outside of the phylum.
450 We also could not clearly delimit Scyphozoa from Hydrozoa, but Hydrozoa were found in all eDNA
451 samples, and therefore have no difference in density. We therefore considered the enrichment test for
452 *putative* Scyphozoa.

453
454 The SMCA was significantly more dense than all other zones in Echinoidea (sea urchins) and
455 Ignavibacteria (a chemoheterotrophic bacteria) and sparser in Perkinsea (parasitic Chromista causing
456 diseases in shellfish and amphibians)(Figure 4; Table S11). The SMCA and tidepools versus other zones
457 were significantly denser in Nitriliruptoria (often obligate halophilic bacteria), denser in
458 Compsopogonophyceae (a red algae), sparser in Bacilli (bacteria including anthrax), sparser in
459 Spirochaetia (ecologically diverse bacteria that includes taxa causing syphilis and Lyme disease) and
460 sparser in Calditrichae (bacteria with few functional attributes known). The SMCA and other beaches
461 harbored higher density of many Animalia including tunicates in Ascidiacea (sea squirts) and
462 Appendicularia, Polyplacophora (chitons), most Protozoa and several algae (e.g. Ulvophyceae,
463 Rhodellophyceae) compared to the marsh area. The marsh was denser in Chromadorea (roundworms),
464 Trematoda (flukes, parasites of mollusks and vertebrates), and Gastropoda (snails and slugs). The
465 richness of complex lifecycle parasites in the marsh suggest high presence of mobile species and
466 genotypic diversity, such is exemplified by high trematode density where snails and migrating birds are
467 present (Hechinger and Lafferty, 2005; Auld and Tinsley, 2015). The marsh area has rich migrating bird
468 diversity, tracked through eBIRD (GBIF.org; see original polygon DOI download for the embayment). The
469 unprotected inner beach, which we note had sites with elevated LCBD scores (Section 3.3), was uniquely
470 dense in Spartobacteria, a class of both soil-dwelling and aquatic bacteria implicated in degrading algal
471 polysaccharides (Herlemann et al., 2013)(Figure 4).

472

473 *3.6. eDNA networks reveal previously undescribed relationships*

474 SpIEC-EASI (SParse Inverse Covariance estimation for Ecological ASSociation Inference) was used to

475 generate community networks within each marker. We analyzed results tables summarized to family

476 that were filtered for a minimum presence of 15% across the 88 sites, which left 201 *18S*, 178 *16S*, and

477 88 *CO1* families. Single marker networks were produced as well as dual marker networks to link *16S* with

478 *18S* and with *CO1*, and these were made into interactive figures for users to explore online (e.g. Figure

479 S5, *Networks* tab). Results produced over 3800 links between families. Two prominent families found to

480 be most integrated in the network were Erythrotrichiaceae (red algae) and Mytilidae (mussels) (Figures

481 S4, Table S12). We queried in the Global Biotic Interactions (GloBI) database, which aggregates

482 published interaction findings, using these two families and 25 others that all had ten or more degrees

483 in the *18S* dataset. We asked how many had published biotic interactions, and if so, how many

484 interactions were consistent. Eleven of those families had no reported interactions in GloBI (Table S13).

485 Six of those families had at least one interaction that was detected in our *18S* network (Table S12). This

486 result supports the notion that eDNA ecological networks can advance inventories of biotic interactions.

487

488 Figure 5 depicts an example of how the interaction networks can be used. The anemone family

489 (Actiniidae) multi-SpIEC-EASI network, determined by *18S* and *16S*, detected interactions with 13

490 families. The GloBI query of Actiniidae retrieved 34 interacting families as either sources or targets, but

491 only one family was also found in our results: Symbiodiniaceae. This symbiotic dinoflagellate has been

492 reported elsewhere as part of the anemone holobiome (León-Palmero et al., 2018; Muller et al., 2018).

493 Our networks also revealed a three-way link among Actiniidae, Symbiodiniaceae, and the bacteria

494 Rubritaleaceae. Weber et al. (2017) published the interactions between Rubritaleaceae and

495 Symbiodiniaceae as part of the coral microbiome. These results demonstrate networks depict candidate
496 holobiomes and microbiomes for larger taxa that can further understanding of ecological functions.
497

498 *3.7. GBIF biodiversity inventories are a different lens of biodiversity*

499 The “Intertidal Biodiversity Survey at Pillar Point” project run by CAS Citizen Science is a focused effort to
500 specifically inventory the low-tide observable biodiversity of the unprotected area tidepools, placing
501 emphasis on tracking certain species such as nudibranchs, and sea stars exposed to sea star wasting
502 disease. Observations in the SMCA or embayment were opportunistic. The survey was sometimes
503 limited when low tide occurred at night. However, the longstanding regularity of citizen science
504 monitoring of Pillar Point (since 2012) has made this beach one of the most iNaturalist observation-rich
505 areas in California. The Survey had 15,539 observations of 596 species on its iNaturalist page when GBIF
506 data were obtained (as of July 5, 2018). Of these observations 12,044 observations were of 457 species
507 and were research grade. We used these data in comparisons with eDNA to understand the differences,
508 parallels, and complementary of the two methods for inventorying and characterizing biodiversity.

509
510 The GBIF observations from the sum of the three polygon downloads were largely iNaturalist in origin,
511 comprising 200 of the 226 ‘protected’ (SMCA) polygon records, 1,697 of the 1,799 ‘embayment’ polygon
512 records, and 11,799 of the 11,900 ‘unprotected’ polygon records (Figure 1; Table S2; 13925 records
513 total). Within the GBIF records, we found 689 species were inventoried with iNaturalist, and 140 species
514 were inventoried by other sources. iNaturalist taxa were mostly members of the kingdom Animalia
515 (78%), but also included Chromista (22%), Fungi (1%), and Plantae (17%). The taxonomic lineages
516 identified in GBIF records totaled 598 Animalia, 23 Chromista, 7 Fungi, 147 Plantae, and a single
517 Protozoan, identified as Mycosphaerellaceae (Choanozoa), that was contributed as a natural history

518 museum collection (Tables S2; Figure 1F). Overall, these results highlight the strength of GBIF data to
519 track the animal kingdom ([GBIF Sources](#)).

520
521 We compared fish in the eDNA *12S* and GBIF results because *12S* targeted vertebrates. The *12S*
522 sequencing results produced a small list, but showed complementarity with GBIF. eDNA *12S* results
523 contained 43 taxa, of which 33 were Teleostei and the remainder were other vertebrates (leopard shark,
524 sea otter, sea lion). Of the fish, 23 were resolved to species, with the remainder resolved to genus or
525 family (Table S7). Silversides (Atherinopsidae) were identified in eDNA but sequences were matched to
526 a genus not known in California, *Odontesthes*. We checked the *12S* reference DNA database, and
527 noticed it only contained four Atherinopsidae genera, and the California grunion genus *Leuresthes* was
528 missing. As of the date of submission of this manuscript, a *12S* sequence for this genus has yet to be
529 published in NCBI (April 2019). The grunion genera recorded in GBIF were all missing from the reference
530 eDNA database, suggesting that gaps in sequencing cause erroneous identification, but GBIF information
531 helped elucidate this problem. In a few cases, eDNA could only identify a taxon to genus, such as
532 *Xiphister*, but of the 71 fish species recorded from Pillar Point in GBIF, there was only a single species in
533 this genus inventoried, the black prickleback *Xiphister atropurpureus*, providing a candidate source
534 species for the DNA signal. Of the species-level eDNA identifications, twelve species were in GBIF with
535 observations from within the Pillar Point area polygons, and eleven species were in GBIF and observed
536 in California but not observed at Pillar Point (Table S14), suggesting that because intertidal observations
537 are limited to low tides and sight range, GBIF fish observations may represent a more narrow
538 spatiotemporal window than eDNA, and eDNA may fill in gaps of which species visit the reef at periods
539 of inundation.

540

541 The most common GBIF observations were families Polyceridae (Mollusca) for Animalia, Laminariaceae
542 (Ochrophyta) for Chromista, Ramalinaceae (Ascomycota) for Fungi, Asteraceae (Thacheophyta) for
543 Plantae. These were not the same as the most frequently observed taxa in eDNA results in those
544 kingdoms: Campanulariidae (Cnidaria) for Animalia, Thraustochytriaceae (Pseudofungi) for Chromista,
545 Glomeraceae (Glomeromycota) for Fungi, and Gigartinaceae (Rhodophyta) for Plantae. These
546 differences in dominant taxa suggest eDNA and GBIF are very different lenses to explore biodiversity.
547 We note that for both datasets, most commonly observed does not equal most common: for
548 photographs, it means more photographs of these taxa have been shared, and for eDNA, it means more
549 samples had DNA from those taxa, and release of DNA into the environment may vary among taxa. The
550 [Common Taxa](#) tab shows which taxa are shared between datasets. We were surprised that the
551 Campanulariidae were never identified in GBIF data from the bay, despite their frequency in iNaturalist
552 observations in other parts of California (GBIF.org; not shown). GBIF records did show observations for a
553 taxon in the same order (Leptothecata), but were all for *Aglaophenia latirostris*. This gap in observations
554 suggests there is capacity to observe broader biodiversity with more overlap with eDNA, but this may
555 require restructuring survey events to generate interest and capacity to notice species.
556
557 GBIF and eDNA analyses detected different aspects of biodiversity, with eDNA detecting 82 phyla and
558 GBIF detecting 19 phyla (Figure 1D). The [Taxonomy Comparison](#) tab allows users to explore the extent of
559 common taxa at the different phylogenetic levels using Venn diagrams (Figure 1E; Figure S6). With
560 increased taxonomic resolution, the concordance between eDNA and GBIF is diminished (Figure 1D).
561 The family level provided the greatest overlap. Only 48 species-level assignments were common to both
562 eDNA and GBIF. Nonetheless, much could be learned from overlapping taxa (see below).
563

564 For example, by comparing frequencies of observation and using the GBIF website (GBIF.org), we were
565 able to develop hypotheses for why the Bay Ghost Shrimp (genus *Neotrypaea*) was observed more by
566 photographs than with eDNA (Figure 1C). Specifically, we hypothesized that the Bay Ghost Shrimp was
567 more populous or widespread at other times of the year when we did not sample for eDNA (i.e. outside
568 of February and April 2017). GBIF records showed the bulk of observations are in May (n=1701) and June
569 (n=918) while fewer than 25 observations were in each of the other calendar months (Figure S7). This
570 suggests that it may be possible that the early year collections of our eDNA sampling was why we did
571 not observe widespread signals. The shrimp may move to deeper sediments in the winter, or recruit
572 each year. We also observed the inverse pattern for the Star Barnacles (family Chthamalidae; Figure 1C).
573 We detected eDNA signals in all zones, but GBIF data only showed them in the unprotected tidepool
574 zone This may be simply explained by limitations to photograph the adults in the rocky sublittoral zone,
575 or alternatively, that during their larval dispersal phase their planktonic microscopic cyprid larvae are
576 wide spread but are impossible to see in the absence of a plankton tow. Barnacles may also be so
577 common that they are not charismatic to photograph, and this is evidence of human preference. We
578 note that barnacles were not species intentionally surveyed by the volunteers coordinated by CAS.

579

580 *3.8. eDNA fills a gap in GBIF-based biodiversity surveys*

581 GBIF data were too sparse in the SMCA to fairly assess alpha and beta diversity (e.g. Figure 1E). Based on
582 eDNA, we expected that the taxon richness should at least mirror the unprotected polygon. The [Area](#)
583 [Diversity](#) Table form shows only 103 unique taxa were in the protected SMCA according to GBIF,
584 compared to 520 in the adjacent unprotected tidepools, which is where iNaturalist CAS citizen science
585 activities were concentrated. In contrast, eDNA results limited to the main kingdoms inventoried in GBIF
586 (Animalia, Chromista, and Plantae) had similar taxon richness for adjacent regions: 1210 taxa in the
587 SMCA and 1546 in the unprotected tidepools (see *Area Diversity* tab).

588

589 We then ask if eDNA may help link compositionally similar sites, and chose to examine sites that
590 clustered within the SMCA samples in eDNA beta diversity ordination plots (Figure 2; Figure S8). The site
591 that consistently clustered tightly and with SMCA samples regardless of marker was PP184-B1 (syn
592 K0184-LB-S1; <https://data.ucedna.com/samples/792>). This single site had ~100 animal taxa, including
593 cabezon, cormorant, mussels, and ostracods. While it was beyond our capacity to follow replicate GBIF
594 work in the SMCA site and test if these sites are similar using other biodiversity metrics, we propose
595 multi-locus compositional similarity as an eDNA approach to find accessible 'surrogate sites' for difficult
596 to access areas.

597

598 *3.9. GBIF taxa and eDNA have numerous singletons*

599 Diminishing overlap with higher taxonomic resolution may be partially explained by misidentifications or
600 rare taxa. Misidentifications can be due to misdiagnosed traits (morphological or DNA sequence). For
601 both eDNA and GBIF, we plotted site frequencies for taxa found, and found datasets were similar across
602 the classification levels for Animalia, the only kingdom with sufficient GBIF data to be fairly comparable
603 (Figure 2). In the eDNA data, 322/742 of species-level taxa were found in only one of the 88 sites (43%
604 singleton species). For GBIF data, 148/566 of species observations were single occurrences (26%
605 singleton species). Although eDNA has significantly more singletons (z-test, z-score -3.6195; $p =$
606 0.00015), singletons occur in over a quarter of species-level inventories in the GBIF dataset. This is a
607 substantial proportion that may indicate the same resolution limitations as eDNA. For example,
608 reference databases may be incomplete for both eDNA and GBIF datasets. We also found the large
609 proportion of singletons in the eDNA results were only apparent for Animalia, Chromista, and Plantae,
610 suggesting that in other kingdoms, the eDNA classification bias is not as large. The online [Taxonomic](#)
611 [Frequency](#) tab gives users the capacity to study which taxa are rare or common in results.

612

613 *3.10. eDNA detects lower trophic levels than photographic observations*

614 We used the GloBI database to explore biotic of interactions of families detected by GBIF and eDNA
615 observations (Table S13). We compared enrichment in a non-directional GloBI category, 'interacts_with',
616 which served as a control, to a directional category, 'eats', which served as the test variable. Chi-Square
617 tests (Table 2) showed the eDNA results contained significantly more taxa that are eaten ($X^2 =$
618 139.72 , $df = 1$, $p\text{-value} < 2.2e-16$) suggesting that eDNA detects lower trophic levels. No differences were
619 found between eDNA and GBIF results for the "interacts_with" term ($p\text{-value} > 0.8$).

620

621 **Discussion**

622

623 We demonstrate that environmental DNA metabarcoding, an emergent citizen and community science
624 tool, adds rich data on ecology that can also be used in biodiversity assessment. However, eDNA analysis
625 is not likely to infer the same communities as human observation or biological collection surveys. eDNA
626 studies such as ours are likely to capture lower trophic levels than photographic observations (Section
627 3.10), and depending on the local environment, may capture different spatiotemporal scales (Section
628 3.1). eDNA may reveal genetic diversity that is as yet unclassifiable and underexplored (Tables S4-S7),
629 and may reveal many novel potential biotic interactions.

630

631 Environmental DNA multilocus metabarcoding can also potentially be used to rapidly characterize the
632 spatiotemporal turnover of a system. However, as a recently developed tool, there are important
633 considerations about sensitivity. Relying on eDNA-based estimates of local beta diversity to identify
634 candidate priority conservation areas (da Silva et al., 2018), for instance, is subject to spurious high
635 scores (Section 3.3). Biotic activities, such as growth, spawning, dispersal, burrowing, or recruitment,

636 may lead to misestimation of the spatial ranges of taxa (Section 3.7). Information sourced from GBIF,
637 scientific literature, or taxonomists, along with sufficient DNA barcoding or genome sequencing, is
638 needed to track misidentification in eDNA datasets (Section 3.4).

639
640 The metabarcoding approach, which amplifies and sequences barcode loci from a mixed sample
641 (Taberlet et al., 2012; 2018), seeks to match DNA reads to reference sequences from voucher specimens
642 to receive a taxonomic assignment (see Cristescu, 2014 for discussion). DNA barcode reference
643 databases still have large gaps across phylogenies (see Curd et al., 2019) that limit discovery of their true
644 taxonomic membership. Moreover, many barcodes lack diagnostic sequence variants for lower
645 taxonomic assignment (Wolfe et al., 1987). These insufficiencies remain despite decades of generating
646 DNA barcodes (beginning in 1982; Nanney, 1982; CBOL et al., 2009), and the formation of consortia
647 designating specific barcodes (Hebert et al., 2003; Hebert and Gregory, 2005; Yao et al., 2010; Schoch et
648 al., 2012) that have given rise to millions of publicly available sequences (~6.7 millions sequences in the
649 Barcode of Life Database (BOLD) as of January 11, 2019; boldsystems.org). As natural areas managers
650 rely on species lists and concrete evidence for inventories, we need to close the data gaps while
651 broadening our strategies for effective monitoring from the taxon to the system community where
652 eDNA can be most valuable as tool for ecological research, monitoring and conservation.

653
654 Raw observational data from GBIF provide a rich but patchy view of biodiversity in space and time,
655 because observations are not made systematically and they are subject to human bias and the
656 technology they use. We are limited by where people can go and by their observational choices and
657 capacities to take photographs or make collections (Section 3.8). The CAS team is currently developing
658 tools to manage these biases. Haphazard sampling for eDNA can have the same bias, but with little
659 effort, systematic sampling events can be organized and executed by volunteers. Accurate identification

660 is also a potential challenge for iNaturalist as it is for eDNA. For photographic observations, machine
661 learning algorithms assist the community in identifications, but the computer vision model is limited to
662 10,000 species currently (https://www.inaturalist.org/pages/computer_vision_demo). Community
663 experts affirm and revise these identifications, but over half of the species identified in iNaturalist have
664 fewer than 20 total observations, as a result, there is little comparative data available to corroborate
665 taxonomic assignments. We found elevated single occurrences of taxa in both eDNA and GBIF datasets,
666 which may partially be explained by misidentification (Section 3.9; Figure 6). The addition of more data
667 will be necessary to distinguish rare from misidentified species, as will integration of different kinds of
668 biodiversity inventories in community platforms such as GBIF. Because of the accessibility of various
669 data sources on the GBIF platform, we were able to show 11 fish species found with eDNA are first
670 reports at the beach, as were 76 phyla that had no record in GBIF, which showcases eDNA as a feasible
671 method to both fill inventory gaps and to include the more challenging branches of the tree of life
672 (Ruppert et al., 2019).

673
674 The pace of change in the science of environmental DNA sequencing and analysis is rapid, which makes
675 the value of environmental sample collections all the more important. Curatorial resources such as the
676 Global Genome Biodiversity Network, the NHMLA Diversity Initiative for the Southern California Ocean
677 (DISCO), and the Genomic Observatories Metadatabase (GeOMe), are helping samples and their
678 associated metadata remain accessible for future research. As techniques for sequencing and analyzing
679 eDNA improve, we may soon readily track intraspecific change or genetic diversity that relates to
680 ecological functions. Projects such as for the Earth BioGenome Project (Lewin et al., 2018) may also use
681 environmental samples for species discovery critical to fill in the tree of life and enhance reference
682 databases. Our brief investigation into Ostracoda genetic variants suggests surveying the unprotected
683 tidepools of Pillar Point are the most likely to have uncataloged species (Appendix 2; Table S8). The

684 value of eDNA-based ecological community networks (Section 3.6; Figure S5), such as the candidate
685 holobiomes (Figure 5) they reveal, may also increase as technology and theory improve.

686

687 The Pillar Point marsh and inner beach had been graded C-F in the most recent Heal the Bay report card
688 (2016). While eDNA did show high density of harmful pathogens (Figure 4), we also find a complex,
689 connected, and extremely diverse environment in both the marsh (highest alpha diversity; dense with
690 complex life cycle parasites) and inner beach (highest LCBD). Our demonstration of eDNA to catalog the
691 rich and dynamic biodiversity of this beach provides exciting evidence that there are many ‘low hanging
692 fruit’ findings pertinent to both management and to basic research in biodiversity that can be gleaned
693 from a few days of fieldwork and <\$4000 USD in DNA library preparation and sequencing costs.

694

695 One of the major conservation questions arising from this case study unique about concerned defining
696 and monitoring the unique characteristics of the SMCA. With eDNA, the SMCA was found to harbor high
697 density of unmonitored groups such as urochordates (Ascidiacea; with known invasive species), marine
698 flatworms (Polycladida), and a unique assemblage of microbial eukaryotes. However, a dearth of
699 photographic data contributions meant many priority taxa were not tracked. Although enhancing the
700 observational data for the SMCA may be possible, our results at least provide a framework for
701 monitoring the system for perturbation without relying on human observation. We suggest that by using
702 surrogate sites identified by low eDNA beta diversity (Section 3.8), responses to environmental stress or
703 change can be measured in heavily surveyed sites, and these can be used to predict the processes
704 occurring in the SMCA.

705

706 The CALeDNA Program is growing its online inventory of California’s biodiversity and participants in the
707 program are encouraged to pair their eDNA sampling with iNaturalist observations. Future directions

708 will include more case studies that compare both tools with additional biodiversity correlates such as
709 those from remote sensing data. We will also work toward better understanding the significance of
710 spatiotemporal trends. It is imperative that CALeDNA and various public and professional biomonitoring
711 programs work together to cross-inform each other so we can make strategic efforts to promote healthy
712 biodiversity.

713 **Data Availability**

714 All raw sequence data are available in the NCBI SRA under project *PENDING ACCEPTANCE*. Software
715 necessary to repeat the analyses are available in DRYAD (<https://doi.org/10.5061/dryad.mf0126f>), which
716 also contains DNA reference databases for *18S*, *12S*, and *16S* markers. The CO1-BOLD reference
717 database is archived in Zenodo (*PENDING*). All other data needed to replicate the results are included in
718 supplemental materials.

719

720 **References**

721

722 Allen, W. (2003). Plant blindness. *BioScience* 53, 926–926.

723

724 Amaral-Zettler, L.A., McCliment, E.A., Ducklow, H.W., and Huse, S.M. (2009). A method for studying
725 protistan diversity using massively parallel sequencing of v9 hypervariable regions of small-subunit
726 ribosomal RNA Genes. *PLOS ONE* doi:[10.1371/journal.pone.0006372](https://doi.org/10.1371/journal.pone.0006372).

727

728 Auld, S.K., and Tinsley, M.C. (2015). The evolutionary ecology of complex lifecycle parasites: linking
729 phenomena with mechanisms. *Heredity* 114, 125–132.

730

731 Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful
732 approach to multiple testing. *Journal of the Royal Statistical Society Series B* 57, 289–300.
733

734 Bird, T.J., Bates, A.E., Lefcheck, J.S., Hill, N.A., Thomson, R.J., Edgar, G.J., Stuart-Smith, R.D.,
735 Wotherspoon, S., Krkosek, M., Stuart-Smith, J.F., et al. (2014). Statistical solutions for error and bias in
736 global citizen science datasets. *Biological Conservation* 173, 144–154.
737

738 Bradley, I.M., Pinto, A.J., and Guest, J.S. (2016). Design and Evaluation of Illumina MiSeq-Compatible,
739 18S rRNA Gene-Specific primers for improved characterization of mixed phototrophic communities.
740 *Applied Environmental Microbiology* 82, 5878–5891.
741

742 Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2:
743 High-resolution sample inference from Illumina amplicon data. *Nature Methods* 13, 581–583.
744

745 Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J.,
746 Fraser, L., Bauer, M., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina
747 HiSeq and MiSeq platforms. *The ISME Journal* 6, 1621–1624.
748

749 CBOL Plant Working Group, Hollingsworth, P.M., Forrest, L.L. et al. (2009). A DNA barcode for land
750 plants. *Proceedings of the National Academy of Sciences* 106,12794–12797.
751 doi:[10.1073/pnas.0905845106](https://doi.org/10.1073/pnas.0905845106).
752

753 Chase, J.M., Kraft, N.J.B., Smith, K.G., Vellend, M., and Inouye, B.D. (2011). Using null models to
754 disentangle variation in community dissimilarity from variation in α -diversity. *Ecosphere*.
755 doi:[10.1890/ES10-00117.1](https://doi.org/10.1890/ES10-00117.1).
756
757 Cristescu, M.E. (2014). From barcoding single individuals to metabarcoding biological communities:
758 towards an integrative approach to the study of global biodiversity. *Trends in Ecology and Evolution* 29,
759 566–571.
760
761 Curd, E.E., Gold, Z., Kandlikar, G., Gomer, J., Ogden, M., O’Connell, T., Pipes, L., Schweizer, T., Rabichow,
762 L., Lin, M., et al. (2019). Anacapa Toolkit: an environmental DNA toolkit for processing multilocus
763 metabarcode datasets. *Methods in Ecology and Evolution*, accepted and in press. *BioRxiv*
764 doi:[10.1101/488627](https://doi.org/10.1101/488627).
765
766 Csardi, G., Nepusz, T. (2006). The Igraph software package for complex network research. *InterJournal*
767 *Complex Systems*, 1695.
768
769 da Silva, P.G., Hernández, M.I.M., and Heino, J. (2018). Disentangling the correlates of species and site
770 contributions to beta diversity in dung beetle assemblages. *Diversity and Distributions* 24, 1674–1686.
771
772 Davis, N.M., Proctor, D.M., Holmes, S.P., Relman, D.A., and Callahan, B.J. (2018). Simple statistical
773 identification and removal of contaminant sequences in marker-gene and metagenomics data.
774 *Microbiome* 6, 226.
775

776 Diaz, S., Settele, J., Brondizio, E. (2019). Summary for policymakers of the global assessment report on
777 biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity
778 and Ecosystem Services. Germany: Secretariat of the Intergovernmental SciencePolicy Platform on
779 Biodiversity and Ecosystem Services. Available from
780 https://www.ipbes.net/sites/default/files/downloads/spm_unedited_advance_for_posting_htn.pdf
781

782 Dray, S., Blanchet, G., Borcard, D. et al. (2016). adespatial: Multivariate multiscale spatial analysis.
783 Retrieved from <https://cran.r-project.org/web/packages/adespatial/index.html>.
784

785 Dunn, O.J. (1964). Multiple comparisons using rank sums. *Technometrics* 6, 241–252.
786

787 Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. (1994). DNA primers for amplification of
788 mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol*
789 *Biotechnol* 3, 294–9.
790

791 Fonseca, V.G., Carvalho, G.R., Sung, W., Johnson, H.F., Power, D.M., Neill, S.P., Packer, M., Blaxter, M.L.,
792 Lamshead, P.J.D., Thomas, W.K., et al. (2010). Second-generation environmental sequencing unmask
793 marine metazoan biodiversity. *Nature Communications*. doi:[10.1038/ncomms1095](https://doi.org/10.1038/ncomms1095).
794

795 Gao, X., Lin, H., Revanna, K., and Dong, Q. (2017). A Bayesian taxonomic classification method for 16S
796 rRNA gene sequences with improved species-level accuracy. *BMC Bioinformatics*. doi:[10.1186/s12859-](https://doi.org/10.1186/s12859-017-1670-4)
797 [017-1670-4](https://doi.org/10.1186/s12859-017-1670-4).
798

799 Groom, Q., Weatherdon, L., and Geijzendorffer, I.R. (2017). Is citizen science an open science in the case
800 of biodiversity observations? *Journal of Applied Ecology* 54, 612–617.

801
802 Guerrero, R., Margulis, L., and Berlanga, M. (2013). Symbiogenesis: the holobiont as a unit of evolution.
803 *International Microbiology* 133–143.

804
805 Hebert Paul D. N., Cywinska Alina, Ball Shelley L., and deWaard Jeremy R. (2003). Biological
806 identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological*
807 *Sciences* 270, 313–321.

808
809 Hebert, P.D.N., and Gregory, T.R. (2005). The promise of DNA barcoding for taxonomy. *Systematic*
810 *Biology* 54, 852–859.

811
812 Hecker, S., Bonney, R., Haklay, M., Hölker, F., Hofer, H., Goebel, C., Gold, M., Makuch, Z., Ponti, M.,
813 Richter, A., et al. (2018). Innovation in Citizen Science – Perspectives on Science-Policy Advances. *Citizen*
814 *Science: Theory and Practice* 3, 4.

815
816 Hechinger, R.F., and Lafferty, K.D. (2005). Host diversity begets parasite diversity: bird final hosts and
817 trematodes in snail intermediate hosts. *Proceedings of the Royal Society B: Biological Sciences* 272,
818 1059–1066.

819
820 Herlemann, D.P.R., Lundin, D., Labrenz, M., Jürgens, K., Zheng, Z., Aspeborg, H., and Andersson, A.F.
821 (2013). Metagenomic de novo assembly of an aquatic representative of the Verrucomicrobial class
822 Spartobacteria. *MBio* 4, e00569-12.

823

824 Hollander, M., Wolfe, D.A., and Chicken, E. (2013). Nonparametric Statistical Methods USA: John Wiley
825 and Sons, Inc.

826

827 Hupało, K., Mamos, T., Wrzesińska, W., and Grabowski, M. (2018). First endemic freshwater *Gammarus*
828 from Crete and its evolutionary history—an integrative taxonomy approach. PeerJ 6.

829

830 Kandlikar, G.S., Gold, Z.J., Cowen, M.C., Meyer, R.S., Freise, A.C., Kraft, N.J.B., Moberg-Parker, J.,
831 Sprague, J., Kushner, D.J., and Curd, E.E. (2018). ranacapa: An R package and Shiny web app to explore
832 environmental DNA data with exploratory statistics and interactive visualizations. F1000Research.
833 doi:[10.12688/f1000research.16680.1](https://doi.org/10.12688/f1000research.16680.1).

834

835 Keeley, J. E., and P. H. Zedler. (1998). Characterization and global distribution of vernal pools. Pages 1–
836 14 in C. W. Witham, E. T. Bauder, D. Belk, W. R. Ferren Jr., and R. Ornduff, editors. Ecology, conservation
837 and management of vernal pool ecosystems—Proceedings from a 1996 conference. California Native
838 Plant Society, Sacramento.

839

840 Kurtz, Z.D., Müller, C.L., Miraldi, E.R., Littman, D.R., Blaser, M.J., and Bonneau, R.A. (2015). Sparse and
841 compositionally robust inference of microbial ecological networks. PLOS Computational Biology.
842 doi:[10.1371/journal.pcbi.1004226](https://doi.org/10.1371/journal.pcbi.1004226).

843

844 Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods 9,
845 357–359.

846

847 Legendre, P., and Cáceres, M.D. (2013). Beta diversity as the variance of community data: dissimilarity
848 coefficients and partitioning. *Ecology Letters* 16, 951–963.
849

850 León-Palmero, E., Joglar, V., Álvarez, P.A., Martín-Platero, A., Llamas, I., and Reche, I. (2018). Diversity
851 and antimicrobial potential in sea anemone and holothurian microbiomes. *PLOS*.
852 doi:[10.1371/journal.pone.0196178](https://doi.org/10.1371/journal.pone.0196178).
853

854 Leray, M., Yang, J.Y., Meyer, C.P., Mills, S.C., Agudelo, N., Ranwez, V., Boehm, J.T., and Machida, R.J.
855 (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for
856 metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers*
857 *in Zoology*. doi:[10.1186/1742-9994-10-34](https://doi.org/10.1186/1742-9994-10-34).
858

859 Lewin, H.A., Robinson, G.E., Kress, W.J., Baker, W.J., Coddington, J., Crandall, K.A., Durbin, R., Edwards,
860 S.V., Forest, F., Gilbert, M.T.P., et al. (2018). Earth BioGenome Project: Sequencing life for the future of
861 life. *PNAS* 115, 4325–4333.
862

863 Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads.
864 *EMBnet.Journal* 17, 10–12.
865

866 McMurdie, P.J., and Holmes, S. (2013). phyloseq: An R package for reproducible interactive analysis and
867 graphics of microbiome census data. *PLOS ONE*. doi:[10.1371/journal.pone.0061217](https://doi.org/10.1371/journal.pone.0061217).
868

869 Meyer, R.S., Curd, E.E., Schweizer, T., Gold, Z., Ruiz Ramos, D., Shirazi, S., Kandlikar, G., Kwan, W.-Y., Lin,
870 M., Friese, A., et al. (2019). The California environmental DNA “CALeDNA” program. BioRxiv.
871 doi:[10.1101/503383](https://doi.org/10.1101/503383).

872
873 Meyer, R., and Drill, S. (2019). Community and citizen science at the UC Division of Agriculture and
874 Natural Resources. Final Report, Center for Community and Citizen Science.
875 https://education.ucdavis.edu/sites/main/files/citizen_and_community_science_at_anr_final_report.pdf
876 [f](#)

877
878 Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J.Y., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka,
879 H., Araki, H., et al. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA
880 from fishes: detection of more than 230 subtropical marine species. Royal Society Open Science.
881 doi:[10.1098/rsos.150088](https://doi.org/10.1098/rsos.150088).

882
883 Muller, E.M., Fine, M., and Ritchie, K.B. (2016). The stable microbiome of inter and sub-tidal anemone
884 species under increasing pCO₂. Scientific Reports 6, 37387. doi:[10.1038/srep37387](https://doi.org/10.1038/srep37387).

885
886 Nanney, D.L. (1982). Genes and phenes in Tetrahymena. BioScience 32, 783–788.

887
888 National Research Council. (2001). Grand challenges in environmental sciences. National Research
889 Council. doi: 10.17226/9975

890
891 O’Donnell, J.L., Kelly, R.P., Shelton, A.O., Samhour, J.F., Lowell, N.C., and Williams, G.D. (2017). Spatial
892 distribution of environmental DNA in a nearshore marine habitat. PeerJ. doi:[10.7717/peerj.3044](https://doi.org/10.7717/peerj.3044).

893

894 Oksanen, J., Blanchet, F., Kindt, R. (2015). Community ecology package ‘vegan’. Retrieved from [https://](https://github.com/vegandevs/vegan)

895 github.com/vegandevs/vegan.

896

897 Patterson, D., Mozzherin, D., Shorthouse, D., and Thessen, A. (2016). Challenges with using names to link

898 digital biodiversity information. *Biodiversity Data Journal* 4. doi:[10.3897/BDJ.4.e8080](https://doi.org/10.3897/BDJ.4.e8080).

899

900 Peters, M.K., Hemp, A., Appelhans, T., Becker, J.N., Behler, C., Classen, A., et al. (2019). Climate–land-use

901 interactions shape tropical mountain biodiversity and ecosystem functions. *Nature* 568, 88–92.

902

903 Poelen, J.H., Simons, J.D., and Mungall, C.J. (2014). Global biotic interactions: An open infrastructure to

904 share and analyze species-interaction datasets. *Ecological Informatics* 24, 148–159.

905

906 Raup, D.M., and Crick, R.E. (1979). Measurement of faunal similarity in paleontology. *Journal of*

907 *Paleontology* 53, 1213–1227.

908

909 Robertson, T., Döring, M., Guralnick, R., Bloom, D., Wieczorek, J., Braak, K., Otegui, J., Russell, L., and

910 Desmet, P. (2014). The GBIF integrated publishing toolkit: facilitating the efficient publishing of

911 biodiversity data on the internet. *PLOS ONE*. doi:[10.1371/journal.pone.0102623](https://doi.org/10.1371/journal.pone.0102623).

912

913 Ruggiero, M.A., Gordon, D.P., Orrell, T.M., Bailly, N., Bourgoin, T., Brusca, R.C., Cavalier-Smith, T., Guiry,

914 M.D., and Kirk, P.M. (2015). A higher level classification of all living organisms. *PLOS ONE* 10.

915 doi:[10.1371/journal.pone.0119248](https://doi.org/10.1371/journal.pone.0119248).

916

917 Ruppert, K.M., Kline, R.J., and Rahman, M.S. (2019). Past, present, and future perspectives of
918 environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and
919 applications of global eDNA. *Global Ecology and Conservation* 17, e00547.
920
921 Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W., and
922 Consortium, F.B. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA
923 barcode marker for Fungi. *PNAS* 109, 6241–6246.
924
925 Taberlet, P., Bonin, A., Zinger, L., and Coissac, E. (2018). *Environmental DNA: For Biodiversity Research*
926 *and Monitoring*. United Kingdom: Oxford University Press.
927
928 Taberlet P, Coissac E, Pompanon F et al. 2012. Towards next-generation biodiversity assessment using
929 DNA metabarcoding. *Molecular Ecology*, 21:2045–2050. doi:[10.1111/j.1365-294X.2012.05470.x](https://doi.org/10.1111/j.1365-294X.2012.05470.x).
930
931 Tipton, L., Müller, C.L., Kurtz, Z.D., Huang, L., Kleerup, E., Morris, A., Bonneau, R., and Ghedin, E. (2018).
932 Fungi stabilize connectivity in the lung and skin microbial ecosystems. *Microbiome*. doi:[10.1186/s40168-](https://doi.org/10.1186/s40168-017-0393-0)
933 [017-0393-0](https://doi.org/10.1186/s40168-017-0393-0).
934
935 Weber, L., DeForce, E., and Apprill, A. (2017). Optimization of DNA extraction for advancing coral
936 microbiota investigations. *Microbiome*. doi:[10.1186/s40168-017-0229-y](https://doi.org/10.1186/s40168-017-0229-y).
937
938 Whittaker, R.H. (1972). Evolution and measurement of species diversity. *Taxon* 21, 213–251.
939
940 Wickham, H. (2009) *ggplot2: elegant graphics for data analysis*. USA: Springer.

941

942 Wolfe, K.H., Li, W.H., and Sharp, P.M. (1987). Rates of nucleotide substitution vary greatly among plant
943 mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences* 84,
944 9054–9058.

945

946 Yao, H., Song, J., Liu, C., Luo, K., Han, J., Li, Y., Pang, X., Xu, H., Zhu, Y., Xiao, P., et al. (2010). Use of ITS2
947 region as the universal DNA barcode for plants and animals. *PLOS ONE*.
948 doi:[10.1371/journal.pone.0013102](https://doi.org/10.1371/journal.pone.0013102).

949

950 Zinger, L., Taberlet, P., Schimann, H., Bonin, A., Boyer, F., Barba, M.D., Gaucher, P., Gielly, L.,
951 Giguët-Covex, C., Iribar, A., et al. (2019a). Body size determines soil community assembly in a tropical
952 forest. *Molecular Ecology* 28, 528–543.

953

954 Zinger, L., Bonin, A., Alsos, I.G., Bálint, M., Bik, H., Boyer, F., Chariton, A.A., Creer, S., Coissac, E., Deagle,
955 B.E., et al. (2019b). DNA metabarcoding—Need for robust experimental designs to draw sound
956 ecological conclusions. *Molecular Ecology* 28, 1857–1862.

957

958 **Legends for Figures, Supplementary Figures, Supplementary Tables**

959

960 **Figure 1.** A-C. Map of the Pillar Point beach with highlighted polygons designating embayment (orange),
961 unprotected (blue), and protected (red) areas used to group sites. A. The 88 eDNA sites are color coded
962 by zones. From top right, marsh runoff (blue), unprotected inner beach (grey), unprotected outer beach
963 (green), unprotected tidepools (red), and protected outer beach (SMCA; pink). B. The 13,942 GBIF
964 observations, color coded by kingdom, with Animalia (blue), Plantae (green), Fungi (pink), and Chromista

965 (red). C. Examples of website functions for taxa contained in both eDNA and GBIF datasets. eDNA sites
966 where a taxon was found are drop pins and GBIF sites where a taxon was found are dots. Bay ghost
967 shrimp are commonly photographed but eDNA only detected them once. Star barnacles were only
968 rarely documented in the unprotected tidepools with photographs, but they are found all over the area
969 by eDNA. D. The total number of unique taxonomic lineages in eDNA, GBIF, or in common. The common
970 taxa are also included in the plotted totals of each separate database. E. Venn diagrams of overlapping
971 and unique taxa found in eDNA and GBIF datasets for the total data (All), for the February 2017
972 collection for eDNA and the month of all February GBIF observations made over all years for which data
973 was available, and for the April 2017 eDNA collection and April GBIF observations made over all years
974 data was available. Colors for polygons are consistent with A and scaled to the relative number of taxa
975 out of the total taxa included in that particular diagram. F. Pie charts of the proportion of unique taxa
976 inventoried by each method. Photo credits: Monterey Bay Aquarium and Wild Kratts Wiki.

977
978 **Figure 2.** Jaccard distance principal coordinate analysis of eDNA results from the markers *16S*, *CO1*, and
979 *18S*, color coded by zone. Results show seasonal change along Axis 1 in *18S* and *16S*, and similar
980 clustering of zones across all three markers.

981
982 **Figure 3.** Relative abundance barplots coloring the most frequently observed 21 species. Grey bars
983 represent the sums of all other taxa. Feb versus April indicate February and April 2017 collection dates.
984 LCBD = Local Contributions to Beta Diversity. MHT = Mean High Tide line, Low or High indicate whether
985 the collection was made below or above the MHT.

986 **Figure 4.** Heatmap of the eDNA-based taxonomic density within classes found in each zone, normalized
987 to the total number of sites observed in each zone. Darker shades indicate higher density. Letters
988 indicate significantly different groupings based on the Dunn post-hoc test, where cells with the same

989 letter are not significantly different from each other, based on corrected p-values. Only classes with
990 significant differences from Kruskal-Wallis test results were plotted. We note Crinoidea were removed
991 because sequences could not be affirmatively identified to be from the class.

992

993 **Figure 5.** Example network of Actiniidae from the 18S and 16S joint co-occurrence network (background)
994 generated with multiSpiecEasi. a = link previously reported in Muller et al. (2018) and GloBI. b = link
995 previously reported in Weber et al. (2017).

996

997 **Figure 6** \log_{10} transformed plot of taxon frequency observed in eDNA or GBIF datasets. All taxa were
998 used if they had been classified minimally to the classification level designated (y-axis). The x-axis is the
999 number of sites where a taxon was observed. Most observations in GBIF are Animalia. Single
1000 observations are high for both eDNA and GBIF at the levels of families through species, most
1001 pronounced in Animalia.

1002

1003 **Figure S1.** Taxon accumulation curves as increasing reads are retained in rarified datasets. Curves were
1004 used to choose rarefaction settings.

1005

1006 **Figure S2.** Alpha diversity plots calculated using species richness, Simpson's index, and the Shannon
1007 index (H') in a. Only Shannon index is shown in b.

1008

1009 **Figure S3.** Relative abundance barplots coloring the top 21 species. Grouping definitions are in Table S1
1010 and Figure S2.

1011

1012 **Figure S4.** Plots to interpret skewness of LCBD scores if holobiomes of DNA swamping taxa are left in or
1013 removed from *18S* results. The dependent variables are binary for either swamping or significantly
1014 different LCBD scores from the rest of their groups. The dependent variables are different
1015 representations of the LCBD score range. A. Plot showing the samples with significant LCBD P-values in
1016 original results experienced either an increase in LCBD or a decrease in LCBD when holobiome-reduced
1017 results were used to recalculate LCBD scores. B. Plot showing the majority of ‘swamped’ samples
1018 experienced a reduction in LCBD when holobiome-reduced results were used to recalculate LCBD. C.
1019 Plot showing the samples that were originally ‘swamped’ now fall into the distribution range of the bulk
1020 of the samples in the holobiome-reduced LCBD results. D. Plot showing the samples that were originally
1021 ‘swamped’ were distributed above the LCBD of the bulk of the samples in the original LCBD results.
1022 Comparing plots D and C demonstrates reduction of the candidate holobiomes of high abundance taxa
1023 helped the samples that experienced swamping look typical.

1024
1025 **Figure S5.** SpieciEasi network analysis results of families detected by eDNA. Top left is a density plot of
1026 the number of degrees taxa have in each network type. SpieciEasi networks are shown with numbers
1027 corresponding to a family with the key in Table S12. Interactive figures of the networks are available on
1028 the Pillar Point website.

1029
1030 **Figure S6.** Degree of overlapping taxa among polygons. Circle sizes are scaled to the number of taxa.

1031
1032 **Figure S7.** Excerpt from the GBIF website of observation frequency over months of the year.

1033
1034 **Figure S8.** Ordination plots as in Figure 2 but with sample names as labels. PP184-B1 is the unprotected
1035 tidepool sample that is closest to the SMCA samples in both *16S* and *18S* datasets.

1036

1037 **Table 1.** Features of the project website, https://data.ucedna.com/research_projects/pillar-point

Website Tab	Function
Intro	Introduces the CALeDNA and GBIF data, and shows the location of the data points on a map of Pillar Point
Occurrence Comparison	Shows the distribution of total presence counts of a taxon in eDNA results or in GBIF observations.
GBIF Sources	Depicts data distribution from iNaturalist vs other contributions
GBIF Taxonomy	Shows the GBIF taxa, whether they are in the NCBI database, and whether they have an eDNA result that matches
Common Taxa	Lists taxa at each classification level that overlap, showing datapoints on the Pillar Point map
Area Diversity	Presents filterable comparisons of unique and common taxa found among the three polygons for eDNA and GBIF dataset
Taxonomy Comparison	Displays Venn diagrams of filterable unique and common taxa between eDNA and GBIF results
Detection Frequency	Presents a filterable sorted list of frequency of occurrences for all taxa in a classification level
Networks	Displays interactive co-occurrence networks at the family level
Biotic Interactions	Presents species frequently observed at Pillar Point with iNaturalist, the known biotic interactions and whether interacting taxa were in GBIF or eDNA datasets

1038

1039

1040 **Table 2.** Number of observations of taxa within GBIF and eDNA datasets included in the Global Biotic

1041 Interactions database, that were used in Chi-Square tests (see main text).

Dataset	Source Eats	Target of Eats	Source Interacts	Target Interacts
GBIF (n=375)	5368 (179 unique)	4690 (262 unique)	3906 (223 unique)	4102 (245 unique)
eDNA (n=1371)	5030 (294 unique)	6091 (444 unique)	4523 (381 unique)	4514 (442 unique)

1042

1043 **Table S1.** eDNA sample metadata

1044

1045 **Table S2.** GBIF data for the three polygons

1046

1047 **Table S3.** Comparison of CO1 settings

1048

1049 **Table S4.** eDNA decontaminated results for *16S* marker

1050

1051 **Table S5.** eDNA decontaminated results for the *18S* marker

1052

1053 **Table S6.** Raw Anacapa output from MiSeq runs and eDNA decontaminated results for the *CO1* marker

1054

1055 **Table S7.** eDNA decontaminated results for the *12S* marker

1056

1057 **Table S8.** Genetic diversity analysis of eDNA results for Ostracods

1058

1059 **Table S9.** ANOSIM results from Raup and Jaccard beta diversity analyses

1060

1061 **Table S10.** Local Contribution to Beta Diversity scores

1062

1063 **Table S11.** Density enrichment test results and statistics

1064

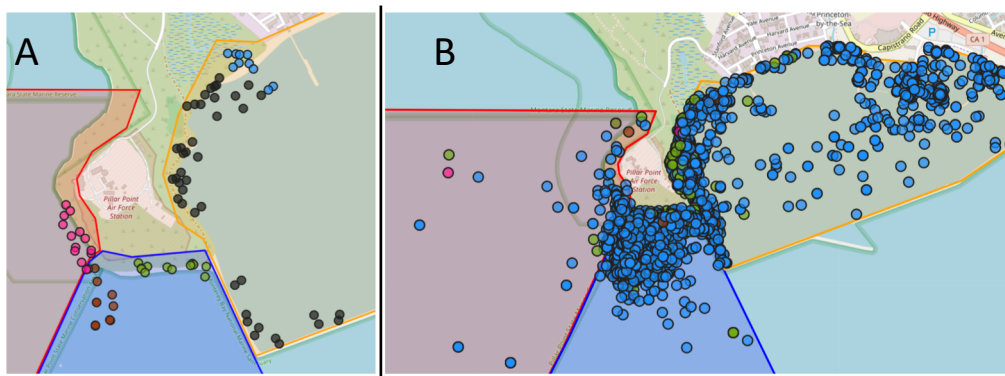
1065 **Table S12.** Family co-occurrence networks

1066

1067 **Table S13.** Global Biotic Interactions data and results

1068

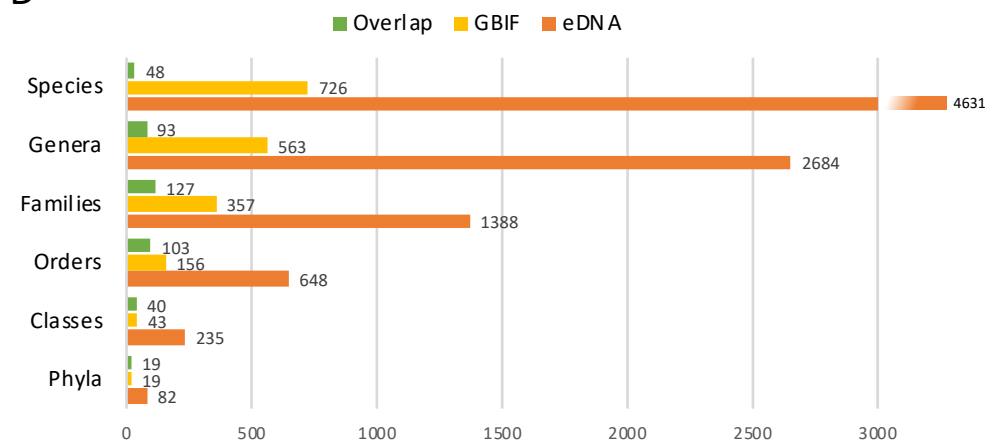
1069



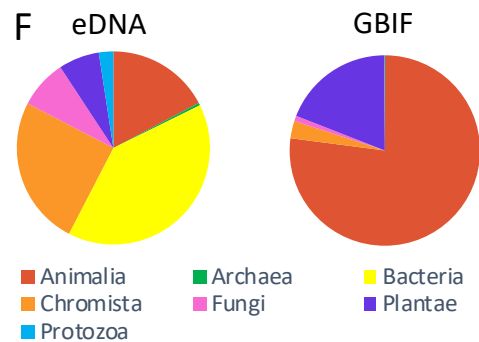
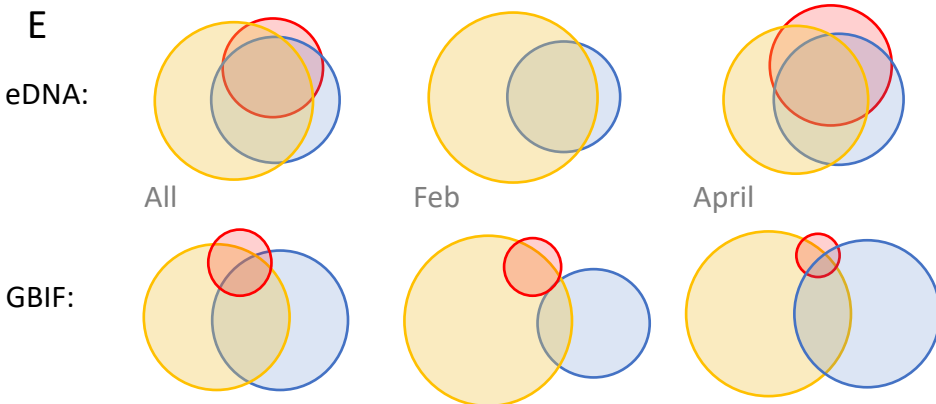
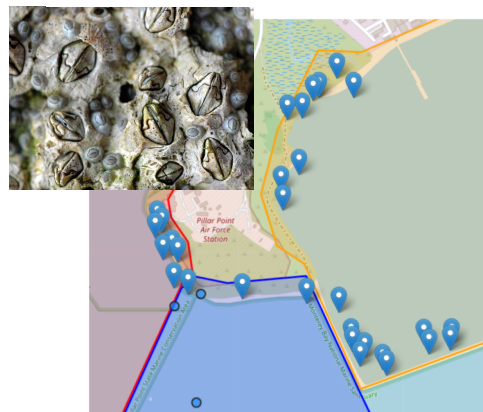
C Genus *Neotrypaea*: Bay Ghost Shrimp



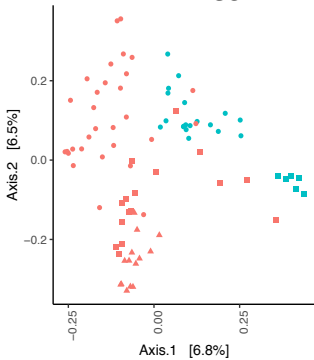
D Taxa reported at Pillar Point by eDNA and GBIF data



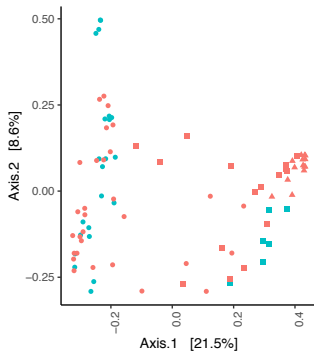
Family Chthamalidae: Star Barnacles



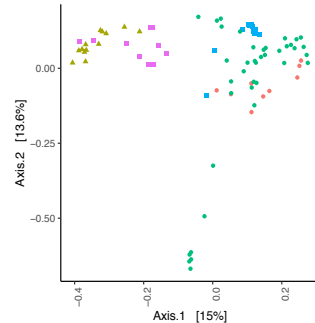
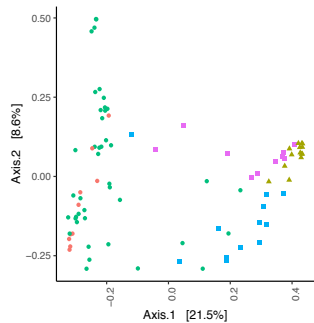
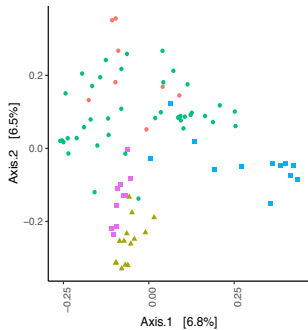
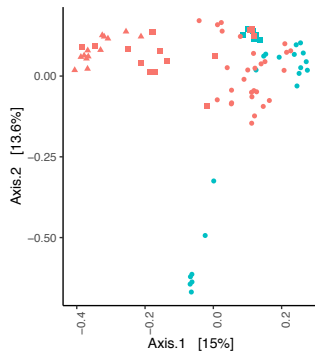
18S

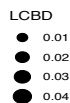
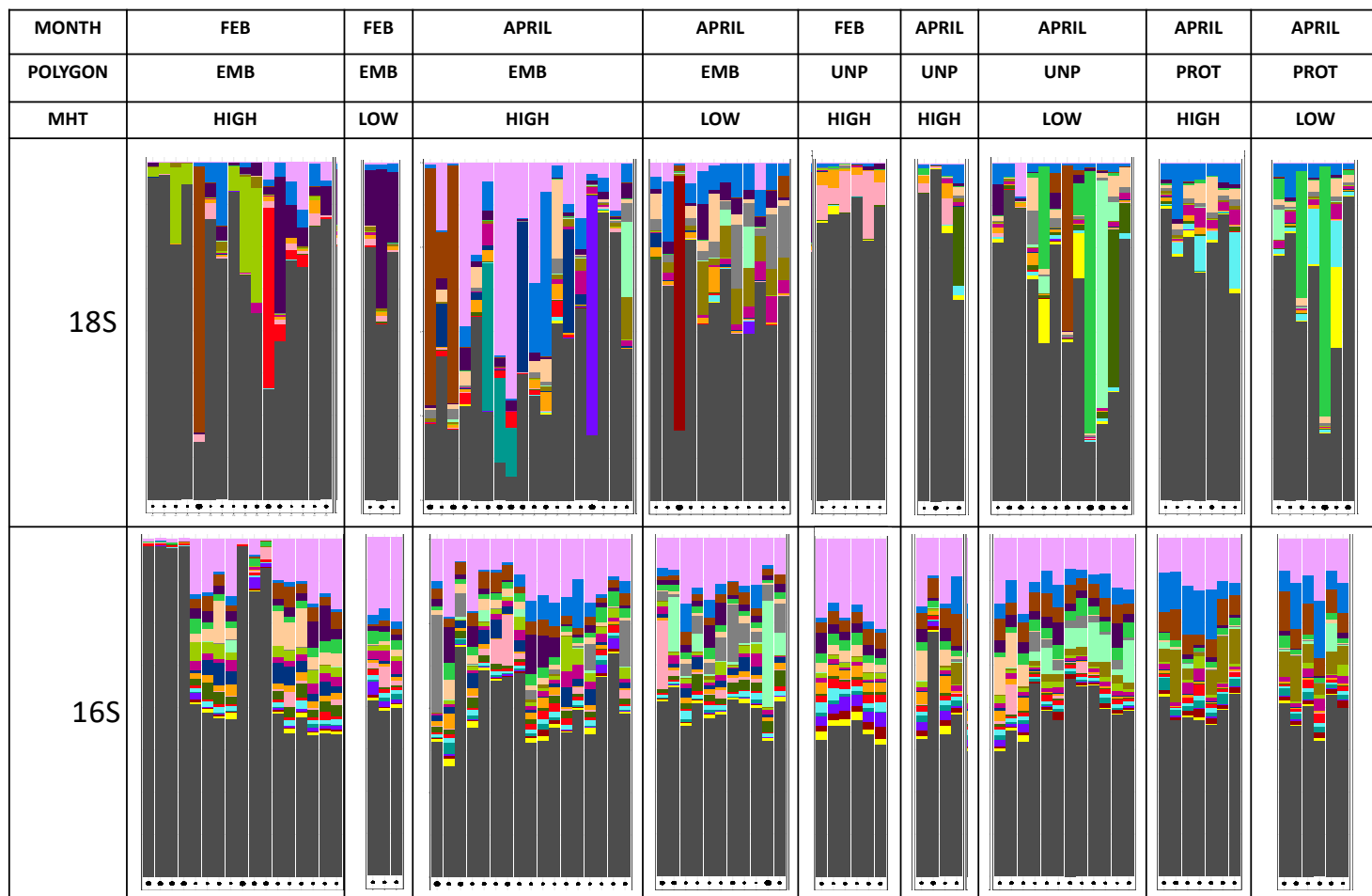


CO1

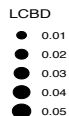


16S





18S



16S

Taxa

- Eukaryota.Miozoa.Dinophyceae.Gymnodiniales.Gymnodiniaceae.Amphidinium.Amphidinium.corpulentum
- Eukaryota.Ochrophyta.Bacillariophyceae.Naviculales.Naviculaceae.Navicula.
- Eukaryota.Nematoda.Enoplea.Enoplida.Oncholaimidae.Pontonema.Pontonema.vulgare
- Eukaryota.Ochrophyta.Bacillariophyceae.Bacillariales.Bacillariaceae..
- Eukaryota.Cnidaria.Anthozoa.Actiniaria.Actiniidae..
- Eukaryota.Bacillariophyta.Coscinodiscophyceae.Thalassiosirales.Thalassiosiraceae.Thalassiosira.
- Eukaryota.Bacillariophyta.Coscinodiscophyceae.Thalassiosirales.Thalassiosiraceae.Minidiscus.
- Eukaryota.Annelida.Polychaeta.Spionida.Spionidae.Rhynchospio.Rhynchospio.glutaea
- Eukaryota.Ochrophyta.Bacillariophyceae....
- Eukaryota.Cercozoa.Thecofilosea.Cryomonadida.Rhogostomidae.Rhogostoma.Rhogostoma.schuessleri
- Eukaryota.Ochrophyta.Bacillariophyceae.Naviculales....
- Eukaryota.Ciliophora.Phylopharyngea.Chlamydomonada.Chlamydomonadidae.Chlamydomonad.
- Eukaryota.Miozoa.Dinophyceae....
- Eukaryota.Cercozoa.Imbricatea.Euglyphida.Ovulinatidae.Ovulinata.Ovulinata.parva
- Eukaryota.Rhodophyta.Florideophyceae.Corallinales.Corallinaceae..
- Eukaryota.Miozoa.Dinophyceae.Gymnodiniales.Gymnodiniaceae..
- Eukaryota.Ochrophyta.Oomycetes.Lagenidiales.Lagenidiaceae.Lagenidium.Lagenidium.giganteum
- Eukaryota.Ciliophora.Nassophorea.Synhymenida.Orthodonellidae.Zosterodasys.Zosterodasys.agamalevi
- Eukaryota.Ciliophora.Nassophorea.Synhymenida.Orthodonellidae..
- Eukaryota.Platyhelminthes.Trematoda.Plagiorchiida.Monorchidae.Monorchis.
- Eukaryota.Rhodophyta.Florideophyceae.Palmariales.Palmariaceae.Palmaria.Palmaria.palmata
- Others

Taxa

- Bacteria.Proteobacteria.Gammaproteobacteria....
- Archaea.Thaumarchaeota.Nitrosopumilales.Nitrosopumilaceae.Nitrosopumilus.
- Bacteria.Bacteroidetes.Flavobacteriia.Flavobacteriales.Flavobacteriaceae..
- Bacteria.Bacteroidetes.Flavobacteriia.Flavobacteriales.Flavobacteriaceae.Maribacter.Maribacter.polysiphoniae
- Bacteria.Proteobacteria.Alphaproteobacteria.Rhodobacterales.Rhodobacteraceae..
- Bacteria.Bacteroidetes.Flavobacteriia.Flavobacteriales.Flavobacteriaceae..Flavobacteriaceae.bacterium.L.1.11
- Bacteria.Bacteroidetes.Flavobacteriia.Flavobacteriales.Flavobacteriaceae.Lutimonas.
- Bacteria.Proteobacteria.Gammaproteobacteria.Vibrionales.Vibrionaceae.Vibrio.
- Bacteria.Proteobacteria.Alphaproteobacteria.Rhodobacterales.Hyphomonadaceae..
- Bacteria.Proteobacteria.Gammaproteobacteria....gamma.proteobacterium.aik8
- Bacteria.Proteobacteria.Gammaproteobacteria....Olavius.algarvensis.Gamma.3.endosymbiont
- Bacteria.Proteobacteria.Deltaproteobacteria.Desulfobacterales.Desulfobulbaceae..
- Bacteria.Proteobacteria.Gammaproteobacteria.Thiotrichales.Thiotrichaceae.Cocleimonas.Cocleimonas.flava
- Eukaryota.Bacillariophyta.....
- Bacteria.Proteobacteria.Deltaproteobacteria....
- Bacteria.Planctomycetes.Planctomycetia.Planctomycetales.Planctomycetaceae..
- Bacteria.Proteobacteria.Gammaproteobacteria.Chromatiales...
- Bacteria.Bacteroidetes.Saprospiria.Saprospirales...
- Bacteria.Proteobacteria.Gammaproteobacteria.Chromatiales.Granulosicoccaceae.Granulosicoccus.
- Bacteria.Actinobacteria.Acidimicrobia....Acidimicrobiidae.bacterium.YM18.244
- Bacteria.Proteobacteria.Gammaproteobacteria.Thiotrichales.Thiotrichaceae..
- Others

Class

Marsh runoff Protected outer beach Unprotected inner beach Unprotected outer beach Unprotected tidepool

Class	Marsh runoff	Protected outer beach	Unprotected inner beach	Unprotected outer beach	Unprotected tidepool		
Polycladidea	ab	a	b	ab	ab	Animalia	
Ascidacea	a	b	a	ab	ab		
Appendicularia	a	b	a	ab	a		
Polyplacophora	a	b	a	b	b		
Mammalia	a	ab	a	b	a		
Echinoidea	ab	c	a	ab	b		
Trematoda	a	b	ab	ab	b		
Anthozoa	a	b	b	b	b		
Demospongiae	a	ab	ab	ab	b		
Ostracoda	ab	ab	a	a	b		
Neophora	a	ab	ab	b	b		
Scyphozoa	a	ab	ab	b	b		
Chromadorea	a	b	a	b	ab		
Gastropoda	a	ab	b	ab	b		
Thermoplasmata	a	b	a	ab	b	Archaea	
Nitrososphaeria	ab	a	b	ab	ab		
Fusobacteria	a	b	a	ab	ab		
Nitriliruptoria	ab	c	a	a	bc		
Halobacteria	a	b	ab	ab	b		
Ignavibacteria	a	b	a	a	a		
Sphingobacteria	a	a	b	ab	a		
Zeta proteobacteria	ab	a	b	ab	a		
Rhodothermia	a	ab	a	a	b		
Bacilli	a	b	a	a	ab		
Spirochaetia	a	b	a	ab	a		
Bacteroidia	ab	a	b	ab	ab		
Spartobacteria	a	a	b	a	a		
Caldilineae	ab	a	b	ab	ab		
Clostridia	a	b	a	ab	a		
Calditrichae	ab	c	a	ab	bc		
Deferribacteres	ab	cd	a	c	bd		
Bicoecae	a	b	c	abc	bc	Chromista	
Centrohelea	a	b	a	a	ab		
Phaeophyceae	a	b	b	b	b		
Polycystinea	ab	a	b	ab	b		
Mediophyceae	a	a	a	b	a		
Giobothalamea	a	a	a	b	a		
Pavlovophyceae	a	b	b	b	b		
Oligohymenophorea	a	ab	b	ab	ab		
Perkinsea	a	b	a	a	a		
Pelagophyceae	ab	a	b	a	a		
Heterotrichae	a	b	a	ab	ab		
Phyllopharyngea	a	a	a	b	a		
Telonemea	a	b	a	ab	ab		
Gregarinasina	a	b	bc	ac	b		
Plagiopylea	a	b	ac	bc	a		
Cryptophyceae	a	b	a	ab	b		
Labyrinthulea	a	b	ac	abc	bc		
Basidiobolomycetes	ab	a	b	a	a	Fungi	
Tremellomycetes	a	b	a	ab	a		
Pyramimonadophyceae	a	bc	b	ac	abc		
Compsopogonophyceae	a	b	a	a	b		
Ulvophyceae	a	b	ab	b	b		
Rhodellophyceae	a	b	a	b	b		
Bangiophyceae	a	ab	b	a	b		
Trebouxiophyceae	a	b	ab	b	ab		
Zygnemophyceae	ab	ac	b	ac	c		
Chlorodendrophyceae	a	b	a	b	b		
Lobosa	a	b	b	ab	b		Plantae
Thecomonadea	a	ab	b	ab	ab		
Choanoflagellata	a	b	b	ab	b		
Ichthyosporea	a	b	b	a	b		
Tubulinea	a	b	b	ab	ab		
							Protozoa

