

1 Message in a Bottle: Archived DNA Reveals Marine Heatwave-Associated Shifts in Fish
2 Assemblages

3
4
5 Zachary Gold^{1,2*}, Ryan P. Kelly³, Andrew Olaf Shelton², Andrew R. Thompson⁴, Kelly D.
6 Goodwin^{4,5}, Ramón Gallego², Kim M. Parsons², Luke R. Thompson^{5,6}, Dovi Kacev⁷, Paul H.
7 Barber⁸
8

9 ¹ Cooperative Institute for Climate, Ocean, & Ecosystem Studies, UW, Seattle, WA

10 ² Northwest Fisheries Science Center, NMFS/NOAA, Seattle, WA

11 ³ School of Marine and Environmental Affairs, UW, Seattle, WA

12 ⁴ Southwest Fisheries Science Center, NMFS/NOAA, La Jolla, CA

13 ⁵ Ocean Chemistry and Ecosystems Division, Atlantic Oceanographic and Meteorological
14 Laboratory, Miami, FL

15 ⁶ Northern Gulf Institute, Mississippi State University, Mississippi State, MS

16 ⁷ Scripps Institution of Oceanography, UCSD, La Jolla

17 ⁸ Department of Ecology and Evolutionary Biology, UCLA, Los Angeles, CA

18

19

20 *Corresponding author Email: zachary.gold@noaa.gov

21

22 Summary: Novel quantitative abundance estimates from archived DNA reveals marine
23 heatwave-associated shifts in fish assemblages.

24 **Abstract**

25 Marine heatwaves can drive large-scale shifts in marine ecosystems, but studying their impacts
26 from a species assemblage perspective can be difficult. Here, we leverage the taxonomic breadth
27 and resolution of DNA sequences derived from the ethanol of a 23-year longitudinal sample
28 collection, combining these with microscopy-derived ichthyoplankton identification to yield
29 higher-resolution species-specific quantitative abundance estimates of fish assemblages in the
30 California Current Large Marine Ecosystem during and after the 2014-16 Pacific marine
31 heatwave. This integrated dataset reveals patterns of tropicalization with increases in southern,
32 mesopelagic species associated and declines in important temperate fisheries targets (e.g. North
33 Pacific Hake (*Merluccius productus*) and Pacific Sardine (*Sardinops sagax*)). We observed novel
34 assemblages of southern, mesopelagic fishes and temperate species (e.g. Northern Anchovy,
35 *Engraulis mordax*) even after the return to average water temperatures. Our novel modeling
36 approach opens the door to reconstructing the historical dynamics of assemblages from modern
37 and archived samples worldwide.

38

39 **Introduction**

40 Climate-induced marine heatwaves (MHWs) are increasing in frequency and severity with far-
41 reaching consequences in marine ecosystems (1) ranging from severe organismal stress to
42 cascading ecosystem effects (2). Notable recent examples include repeated bleaching events
43 across the Great Barrier Reef (2016, 2017, 2020) (3) and near-total kelp deforestation in
44 Northern California, USA (2016-19)(4). These marine heatwaves precipitated drastic,
45 unprecedented changes in dominant foundational species across hundreds of thousands of square
46 kilometers of shallow, coastal ecosystems.

47 The impacts of such large environmentally driven disturbances on coastal marine
48 ecosystems have been ecologically and economically significant (5–7). In the 1940s, the
49 dramatic collapse of Pacific Sardine (*Sardinops sagax*) disrupted marine foodwebs, causing
50 broad-scale, negative socio-economic impacts across the Northeast Pacific (8–10). To better
51 understand the processes driving these complex marine ecosystem dynamics and to avert similar
52 fisheries collapses within the California Current Large Marine Ecosystem (CCLME), the
53 California Cooperative Oceanic Fisheries Investigations (CalCOFI) was formed in 1949.
54 CalCOFI has continuously conducted systematic fisheries-independent surveys of the southern
55 CCLME from 1951 until present (11, 12), focusing on monitoring larval fish assemblages, as
56 larval fish dynamics are a key predictor of ecosystem health and function (5, 12, 13).

57 Larval fish abundances help to characterize the state of marine ecosystems as they track
58 spawning-stock biomass (14). Over 70 years of CalCOFI research has documented decadal and
59 annual changes in fish assemblages in response to environmental conditions, identifying major
60 shifts in response to Pacific Decadal Oscillations and El Niño Southern Oscillations (12, 15–17).
61 These decadal and annual changes in ichthyoplankton dynamics are superimposed over the
62 strong biogeographic assemblage associations with distinct water mass characteristics within the
63 Southern California Bight (17). Ichthyoplankton assemblages differ among the colder and fresher
64 California Current, warmer and saltier California Counter Current and Central Pacific water
65 mass, and in upwelling conditions across the continental shelf (11, 13, 18, 19). Importantly,
66 periods of elevated temperatures were historically associated with higher abundances of
67 southern, mesopelagic species and Pacific Sardine while colder periods were associated with
68 higher abundances of northern, mesopelagics and Northern Anchovy (*Engraulis mordax*) (10,

69 20). These insights into forage fish community dynamics across decadal climatic regime shifts
70 are vital to understanding the effects of climate change on the CCLME (9, 11, 19).

71 Despite the value of previous CalCOFI ichthyoplankton efforts, such traditional manual
72 identification of larvae is labor-intensive, and taxonomic resolution is often limited by a lack of
73 discernible morphological characteristics (21). Here, we reconstruct ichthyoplankton
74 assemblages over 23 years, applying genetic metabarcoding (amplicon sequencing) to the
75 ethanol surrounding preserved CalCOFI plankton samples. We pair these data with
76 morphological count observations in a joint Bayesian model to estimate species-specific larval
77 abundance.

78 We model taxon sequence-read counts from metabarcoding as an outcome of the PCR
79 process, wherein each taxon has a different amplification efficiency for the primer set used (see
80 Materials and Methods). We link the sequencing data to morphological ichthyoplankton counts
81 to constrain the species-specific starting concentrations of DNA in the ethanol jars. The resulting
82 integrated model leverages the taxonomic breadth and resolution (22, 23) of amplicon
83 sequencing, combining these with the power of morphological counts to yield species-specific
84 quantitative abundance estimates. By jointly interpreting these datasets, we can track changes in
85 abundance for a broad diversity of larval fish species, yielding a much higher-resolution picture
86 of these assemblages.

87 The resulting estimates capture major and sometimes highly unexpected changes to the
88 fish assemblages during and after the 2014–2016 Pacific Marine Heatwave (MHW), the warmest
89 3-year period in the North Pacific over 100 years of recorded history (24). We directly examine
90 the responses of ichthyoplankton assemblages to sea surface temperature (SST) as an abiotic
91 proxy for the MHW as a whole, recognizing that a suite of environmental variables including

92 upwelling strength and location, dissolved oxygen, and salinity, among other environmental
93 covariates, changed dramatically during the climate change influenced MHW (25–28).
94 Complementary analyses using mean water-column temperature – rather than SST – yielded
95 similar results (See Supplement S1).

96

97 Results

98 Metabarcoding with MiFish *12S* (29) generated a total of 59.9 million sequence reads
99 across 84 jars representing 90 unique DNA extractions and 262 unique PCR technical replicates.
100 All sequence data were processed using the *Anacapa Toolkit* (30). After quality control,
101 sequence-variant (ASV) dereplication, and decontamination processes (22, 30, 31), we retained a
102 total of 54.5 million reads (See Supplement 1 Methods). From these data, we classified 130
103 unique taxa including 103 species-level assignments (79%), 15 genus-level assignments (12%),
104 11 family-level assignments (8.5%), and 1 class-level assignment. We identified two distinct
105 morphologically indistinguishable lineages of the Northern Lanternfish (*Stenobrachius*
106 *leucopsarus*). The two lanternfish lineages exhibited dramatically different ecological patterns
107 across the samples and were therefore treated separately.

108 Independent microscopy-count data from paired, matching formalin-preserved samples
109 consisted of 9,610 larvae sorted across 84 jars. From these data, we classified a total of 92 unique
110 taxa including 76 species-level assignments (83%) and 16 genus-level assignments (17%).

111 For our integrated Bayesian model, we focused on the 56 species that had sufficient
112 representation across the metabarcoding data set to achieve model convergence (observed in >10
113 technical PCR replicates) and thus provided reliable quantitative estimates (Figures S3–S5).

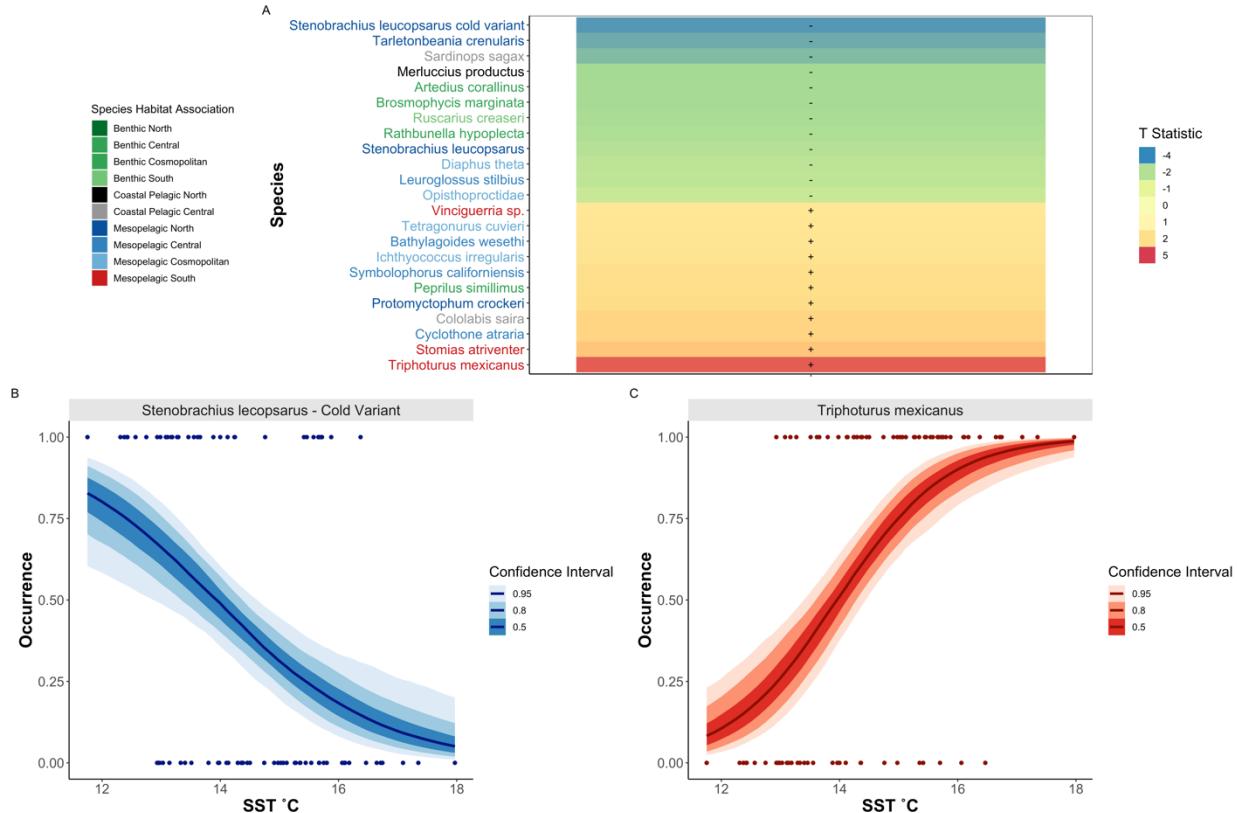
114 Model fits yielded station-, species-, and year-specific larval abundances for 56 fish species
115 spanning a 23-year period.

116 *Displacement of Target Fish Species and Tropicalization of Fish Assemblages Associated with*
117 *the Marine Heatwave*

118 We observed a transformation of marine ichthyoplankton assemblages during the 2014–
119 2016 MHW where southern, mesopelagic species increased while several temperate species of
120 critical ecological and economic importance declined. Such synchronous changes in the marine
121 ichthyoplankton assemblages occurred during the MHW despite the hundreds of kilometers
122 between stations and unique biogeographic characteristics associated with each sampled
123 geographic location (See Supplement 1 results). For example, the mesopelagic Mexican
124 Lampfish (*Triphoturus mexicanus*) was at peak abundance during the MHW, and extended its
125 typical range both poleward and into coastal shelf waters (Figures 1a & c, S6-10).

126 Shifts were observed throughout the ichthyoplankton assemblages of the study region
127 (Figures 2-3). On the one hand, subtropical, mesopelagic species uniformly increased in
128 association with elevated sea-surface temperatures. On the other hand, many coastal species
129 typically seen in the region tended to decrease (Figure 2, S6-26). In particular, the abundances of
130 northern, mesopelagic species and fisheries targets such as Pacific Sardine (*Sardinops sagax*)
131 and North Pacific Hake (*Merluccius productus*) were significantly correlated with colder sea
132 surface temperatures and displayed negative co-occurrence patterns with warm associated
133 southern, mesopelagic taxa (Figure 3, S27-28).

134



135

136

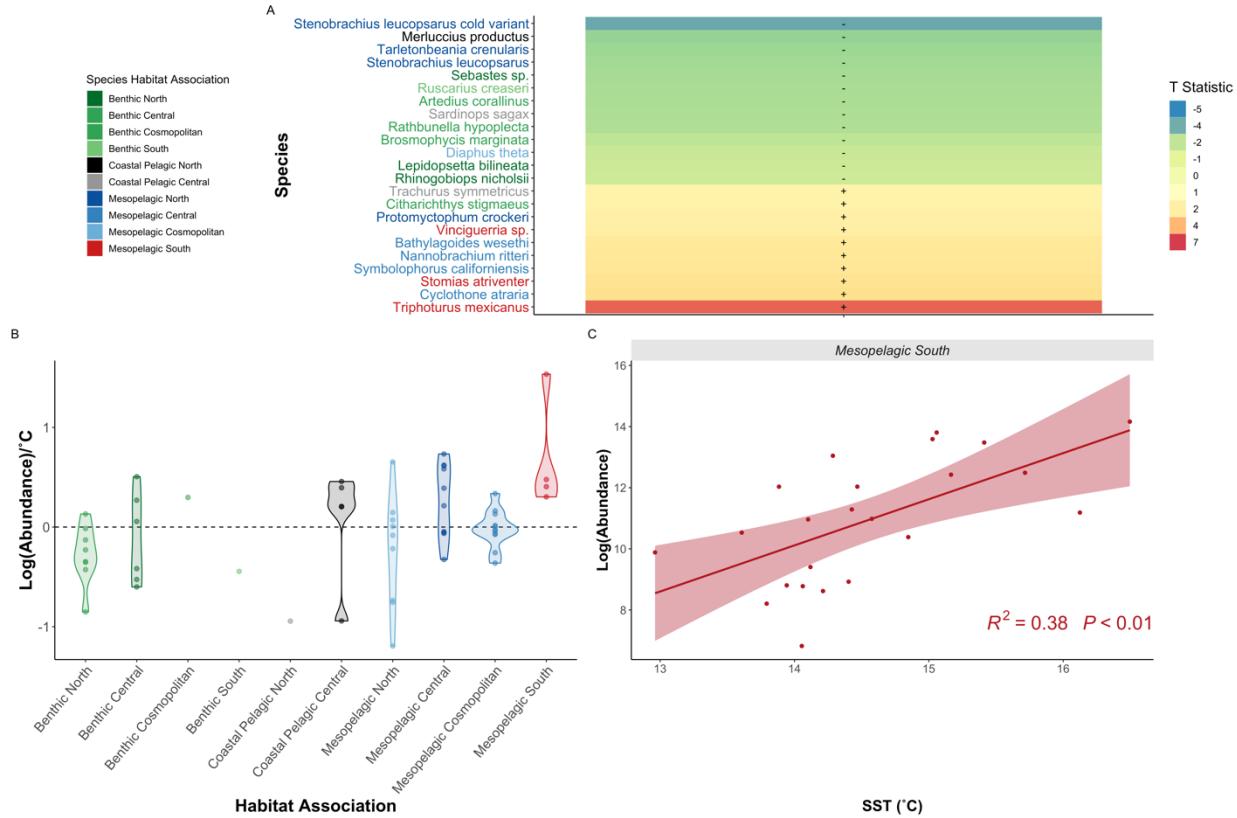
137

Figure 1. Temperature-Occurrence Associations in Fish Species

Changes in species occurrence patterns associated with SST, with southern, mesopelagic species increasing in prevalence with elevated temperature (A). T statistic (slope coefficient / standard error) from a generalized binomial mixed model was calculated for each species across all stations. Only 23 out of 56 total species with significant slopes (95% CI that exclude zero) are plotted. Importantly, metabarcoding identified cold-associated variants of the Northern Lanternfish (*Stennobrachius leucopsarus*) that cannot be morphologically identified (B) as well as common warm-associated species such as the Mexican Lampfish (*Triphoturus mexicanus*) (C).

146

147

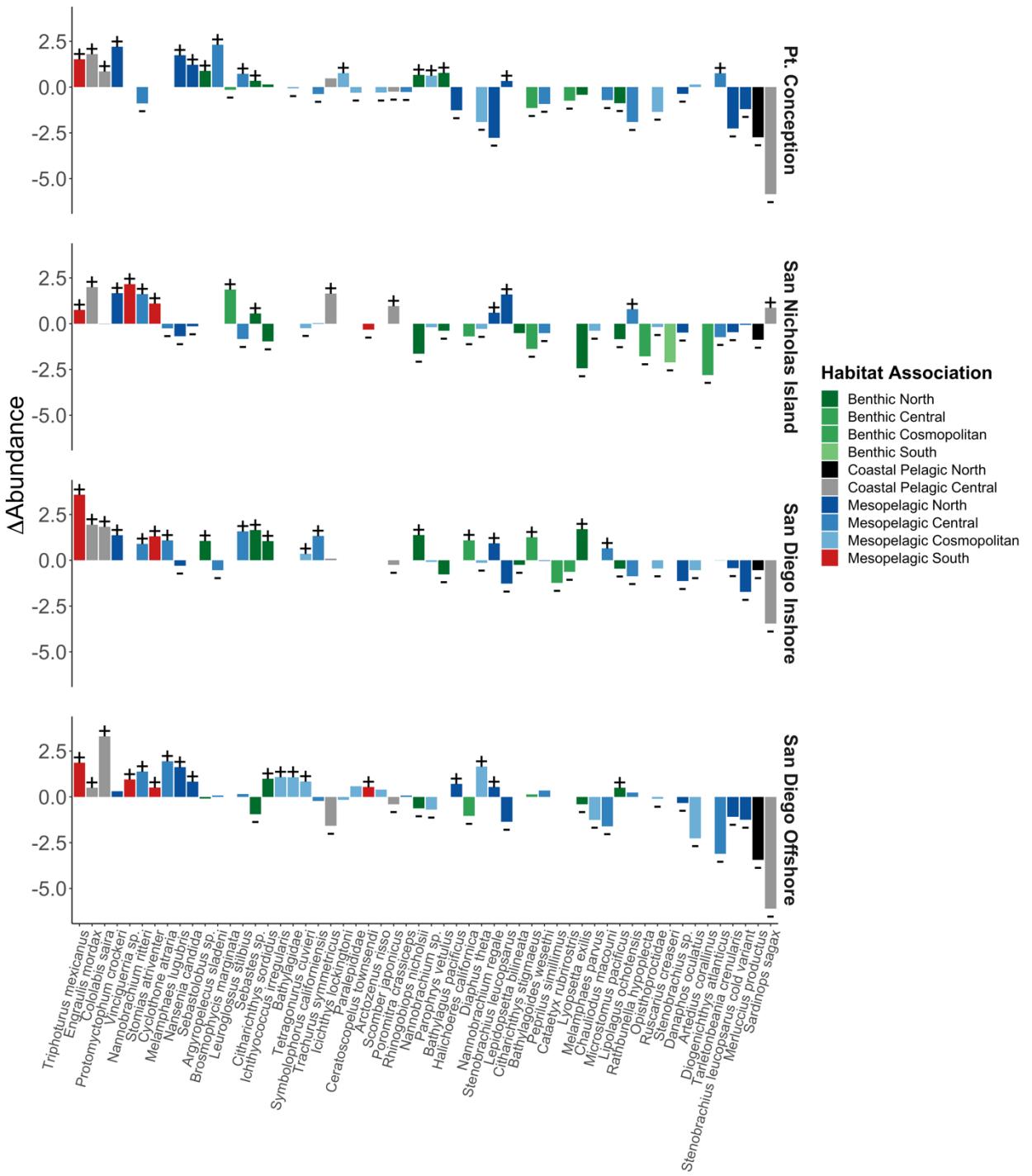


148

149 **Figure 2. Increased Abundance of Southern Oceanic Species Drive Fish Community
Shifts**

151 Changes in species abundance in response to SST °C, with southern, mesopelagic species
152 increasing in abundance with elevated sea surface temperature (A). As in Fig 1,
153 significant T statistics from generalized linear models are plotted. Southern mesopelagic
154 fishes were associated with increased temperature as indicated by summarizing species-
155 specific slopes from generalized linear models (B) and by the aggregated abundance
156 relationship (C; each point is the mean posterior estimate of abundance for all southern,
157 mesopelagic species in a single year). In contrast, abundances of benthic species and
158 coastal pelagic species such as Northern Hake and Pacific Sardine abundances, were
159 associated with cooler temperatures.

160



161

162

Figure 3. Novel Marine Heatwave Assemblages

163

Shifts in species abundances before the MHW (1996-2013) compared to during and after

164

the MHW (2014-2019). Synchronous increases in southern, mesopelagics and Northern

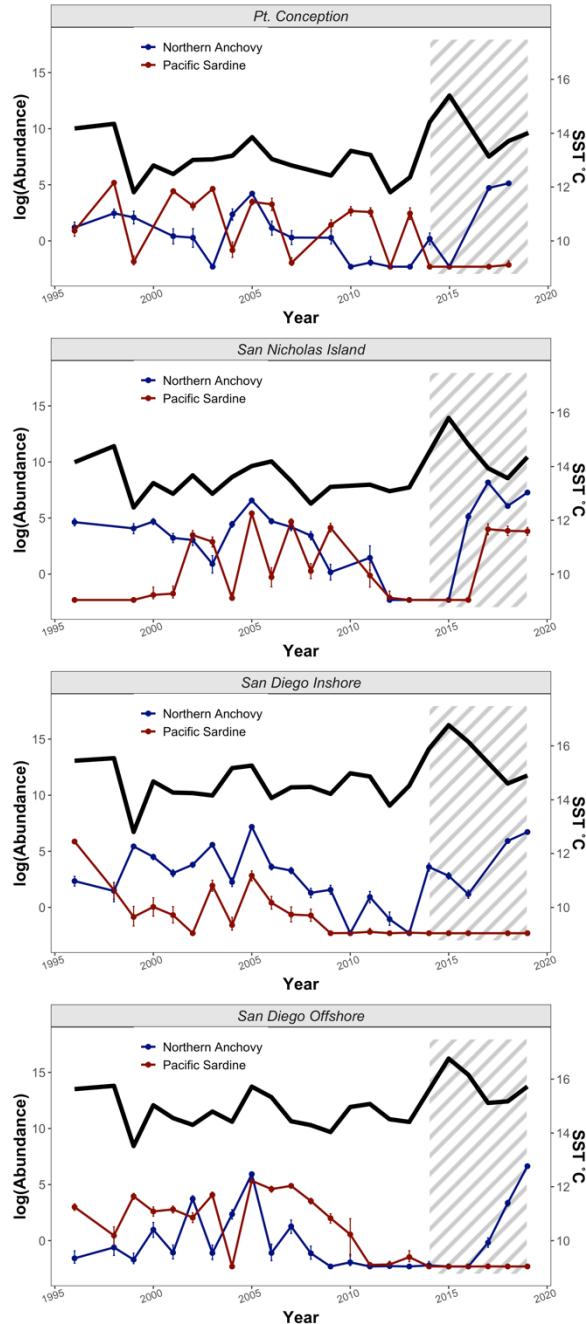
165

Anchovy were observed across all stations. Here, stations are shown in rows, species in

166 columns, and the change in abundance between the two ecological phases shown as the
167 response variable. Fisheries targets including Pacific Sardine and North Pacific Hake, as
168 well as many other benthic and coastal species, had concurrent negative associations.
169 Significant differences during and after the MHW are marked with + or -.

170 *Biomass Changes in Forage Fishes*

171 Both because of their own commercial value and because they are prey for other high-
172 value fishery species, sardine and anchovy fluctuations have been a major focus of fisheries
173 research since the 1950s (9). Our model estimates are consistent with other studies (20, 32) that
174 documented a decline in both sardines and anchovy beginning in 2005. This period of time was
175 followed by high abundances of anchovy, but continued low abundances of sardine in the wake
176 of the MHW, 2015-2019 (Figure 4). Although anchovy larvae abundance was low in spring
177 during the 2014-2016 MHW, anchovy recruitment was high in summer 2015 (33). Anchovy
178 mature in approximately one to two years (32, 34) and thus the 2015 class likely began spawning
179 in mid-2016 (20), leading to high anchovy spawning stock biomass and larval abundances by
180 2016 and lasting into 2019 and 2020 (35).



181

182 **Figure 4. Synchronous Increase in Anchovy Abundance During and After Marine
183 Heatwave**

184 Posterior estimates for larval fish abundances (counts/10m²) over time at each of the four
185 sampled stations. Joint modeling of metabarcoding and morphological counts
186 reconstructed increases in Northern Anchovy (*Engraulis mordax*) [blue] during the recent

187 Pacific Marine Heatwave and low spawning of Pacific Sardine (*Sardinops sagax*) [red]
188 over the past decade (points are means and error bars are 95% credible intervals; shaded
189 region is during and after the MHW). SST is plotted above the Northern Anchovy and
190 Pacific Sardine abundances, for reference.

191

192 **Discussion**

193 *Displacement of Target Fish Species and Tropicalization of Fish Assemblages Associated with*
194 *the Marine Heatwave*

195 Recent studies demonstrate the tropicalization of terrestrial and marine ecosystems in
196 response to climate change (40, 41). These shifts can induce novel species interactions,
197 catalyzing changes in ecosystem function (2, 25). For example, we observe the combination of
198 high abundances of both Northern Anchovy and southern, mesopelagics (5, 38) – a pattern
199 otherwise undocumented in the previous >70-year CalCOFI dataset (16, 20). Furthermore, our
200 results suggest multiple coastal pelagic fisheries targets may continue to be scarce as
201 environmental conditions that are similar to the 2014-2016 MHW become more common (5, 38,
202 39). Although the ecological implications of these novel assemblages are, by definition,
203 unpredictable, our results suggest that if future assemblages resemble those seen in the MHW,
204 increases in Northern Anchovy and southern, mesopelagic fishes are likely to be associated with
205 decreases in Pacific Sardine and North Pacific Hake in the Southern CCLME (40, 41); these
206 shifts have fundamentally changed ecosystems and fisheries relative to the recent past (20).

207

208 *Biomass Changes in Forage Fishes*

209 The rise in anchovy and continued low abundances of sardine during the MHW is an
210 ecological surprise. Correlative analyses between basin-scale environmental indices such as the
211 Pacific Decadal Oscillation indicate that, for the latter half of the 20th century, anchovy thrived
212 under cooler conditions and sardine under warmer conditions (10). However, our findings and
213 others (5, 9, 38, 42) suggest that the mechanisms that govern the population dynamics of these
214 species are not a mere function of temperature, but that more complex factors drive recruitment
215 dynamics of these species (20). For example, despite largely synchronous responses of fish
216 assemblages to the marine heatwave, sardine declines were not consistent across the CCLME,
217 with refugia of localized abundance in nearshore waters potentially driven by distinct, favorable
218 conditions (9, 20, 32).

219 Further improving our mechanistic understanding of drivers of fish assemblage dynamics
220 will better inform ecological predictions in the face of extreme ocean events such as MHWs
221 which are likely to increase in frequency and duration under climate change (43). As we
222 demonstrate, a combination of metabarcoding and visual surveys can characterize species across
223 trophic levels (44) and this has the potential to reveal ecological mechanisms. Here, we used
224 metabarcoding to accurately characterize the composition of larval fishes in CalCOFI plankton
225 samples. Future efforts could focus on documenting the phytoplankton and zooplankton
226 assemblages that comprise both larval prey and predators. Several major hypotheses seeking to
227 explain recruitment variability are underpinned by the capacity of young larvae to consume
228 appropriate prey that facilitates faster growth (45). Unfortunately, accurately characterizing the
229 larval prey field has traditionally been difficult as prey are generally too small to be accurately
230 sampled by nets (46). Metabarcoding of water samples from the same locations where larvae are
231 collected, however, can characterize the larval prey field at an unprecedentedly high level of

232 detail. In addition, metabarcoding of the stomachs of larval fishes can then identify actual prey
233 items that were consumed by larvae. Evaluating the larval prey field and gut contents through
234 metabarcoding will help us to finally understand the drivers of recruitment volatility in coastal
235 pelagic and other fishes (5, 32, 47–51).

236 The described unexpected rise in anchovy during the 2014–16 MHW resonated
237 throughout the CCLME (39). For example, California sea lion pups grew at anomalously high
238 rates after their mothers consumed copious anchovy forage and produced ample milk (40). High
239 rates of almost exclusively anchovy consumption also seemingly induced thiamine deficiency in
240 adult salmon resulting in poor condition of recruits (52). Birds capable of feeding on anchovy
241 thrived (38) while those unable to consume anchovy perished (41). Given that conditions
242 comparable to the 2014–2016 MHW are predicted to be more common in the CCLME in the
243 future (1), our results suggest that continued biological responses to both anchovy-dominated
244 forage-fish assemblages and MHW-associated ocean warming conditions are likely to be without
245 modern analog (20).

246 *Novel Insights from Legacy Collections*

247 Molecular tools differentiated variants and species that were not morphologically identifiable in
248 the ichthyoplankton (21). For example, metabarcoding identified unique cold-associated variants
249 of the Northern Lampfish (*Stennibrachius leucopsarus*) that are morphologically
250 indistinguishable and combined as a complex are only weakly associated with cooler
251 temperatures. By illuminating such unseen variation, molecular methods reveal ecological
252 dynamics otherwise hidden by shared larval morphology.

253 Ultimately, our approach to studying historical fluctuations in ichthyoplankton
254 assemblages reveals climate-associated biological changes in the CCLME, and suggests ways in

255 which these changes could alter the function and socio-economic benefits derived from marine
256 ecosystems. Importantly, this novel approach provides quantitative estimates by non-
257 destructively sampling legacy collections via metabarcoding, and at the same time provides a
258 mechanistic framework for determining absolute abundance estimates from compositional
259 amplicon sequencing data (53–56). Here we relied on morphological counts to ground
260 compositional metabarcoding data, but our framework suggests that any estimate of abundance
261 (e.g. qPCR) or amplification efficiency (e.g. derived from mock communities) can achieve
262 similar results (See Supplement 2) (55, 56). Unlocking such quantitative metabarcoding
263 approaches expands the potential for linking ecological assemblages to environmental processes
264 beyond just presence-absence analyses (57–59). Such quantitative approaches may prove critical
265 in modeling and predicting future ecosystem change, although directly linking assemblage
266 dynamic responses to climate-driven forces remains inherently challenging. While the CalCOFI
267 samples are specific to ichthyoplankton from the CCLME, bulk collection of community
268 samples is commonly used to survey plankton, insects, pollen, gut contents, and microbiomes,
269 among many other targets (60). As such, here we provide broadly applicable methodology with
270 which to efficiently understand modern and historical changes in ecological communities.

271

272 **Materials and Methods**

273 *Study Design*

274 To investigate decadal changes in the ichthyoplankton assemblages in the California Current
275 vicinity, we identified ichthyoplankton from four distinct stations during spring months.
276 Archived spring ichthyoplankton samples were collected across four biogeographically
277 dissimilar stations with variable water properties (61) over 2 decades (1996, 1998–2019; Figure

278 S1; See Supplement 1 methods)(5, 38). Decades of research within the study region (16, 17, 20,
279 62) indicate the majority of species spawn in spring and closely track adult biomass (63). Hence,
280 we expect the spring ichthyoplankton to reflect underlying changes in the local fish assemblages.

281 At each station, oblique bongo net tows were conducted from 210 m depth to the surface
282 using standard CalCOFI methods (15, 64–66). Cod end contents of both bongo nets were
283 preserved at sea. The starboard side was preserved in sodium borate-buffered 2% formaldehyde
284 and the port side was preserved in Tris-buffered 95% ethanol. Microscopy was conducted to
285 identify species abundance from formaldehyde-preserved samples following standardized
286 CalCOFI techniques (61). DNA metabarcoding was conducted on the ethanol in which port side
287 samples were preserved using the MiFish Universal Teleost (67) PCR primer set targeting the
288 12S rRNA mitochondrial gene region. See Supplement 1 methods for full description.

289 *Estimating Abundance*

290 We estimated the abundance of ichthyoplankton in each jar using a novel joint Bayesian
291 hierarchical model described in detail in Supplement 2. In brief, we estimate that the number of
292 sequenced amplicons, for any species i , is proportional to the species-specific fraction of DNA in
293 the PCR template (53, 55, 56). The amplicons produced during a PCR reaction are dictated by
294 the amplicon efficiency parameter a_i , which is characteristic of the interaction between the
295 particular primer set and each species being amplified. Thus, for any species i , the number of
296 amplicons should be directly related to the efficiency of amplification and the starting
297 concentration of DNA template such that

298
$$A_i = c_i(a_i + 1)^{N_{PCR}} \quad (1)$$

299 where A_i is amplicon abundance, c_i is the true number of DNA copies in the reaction attributable
300 to species i , a_i is the species-specific amplification efficiency (bounded on (0,1)), and N_{PCR} is
301 the number of PCR cycles used in the reaction (68).

302 However, not all molecules of DNA are transferred through each molecular step (69, 70),
303 changing what in eq. 1 appears to be a single-species process – each species being amplified
304 independently – into a multi-species process; the number of amplicons observed for species i will
305 depend upon both the amplicons produced for species $i = 1$ and the amplicons produced for
306 species $i = 2, 3, \dots, I$ in the same reaction. Observations of amplicons are thus compositional
307 data, meaning they are the proportions of the sample amplicon reads and convey relative
308 quantitative information of the observed species and therefore need to be analyzed as such (54).

309 To retain the data-generating structure from eq.1 as much as possible, we develop a
310 model for a single sample with many species. As above, if we let I index species with $I = 1, 2, \dots,$
311 I , then we can write a deterministic equation for the number of amplicons observed in log-space
312 as

313
$$\log(A_i) = \log(c_i) + N_{PCR} \log(a_i + 1) + \log(\eta) \quad (2)$$

314 where the only new term is η , representing the proportion of reads observed from a given
315 sampling run (See supplement 2 for details). Note that in this formulation η is a single value
316 shared across all species and serves to scale the number of amplicons observed. Additionally, we
317 can rewrite the number of DNA copies in terms of proportional number of counts, $\log(\beta_i) =$
318 $\log(c_i) - \log(\sum_i c_i)$. Note that the second term in this equation is a sum of the counts across all
319 species, and so is a single shared value for all species. As such it can be integrated into the value
320 η that scales the overall abundance for each species i ,

321
$$\log(A_i) = \log(\beta_i) + N_{PCR} \log(a_i + 1) + \log(\eta) \quad (3)$$

322 This equation is appealing because it provides a process-oriented description of the
323 biology of interest (the β s), a term for how PCR modifies our amplicon sequence count
324 observation of the true abundance ($N_{PCR} \log(a_i + 1)$), and a term for how subsampling of DNA
325 reads will adjust the number of amplicons observed ($\log(\eta)$).

326 However, in the absence of additional information, it is not possible to estimate the β and
327 a parameters from metabarcoding data alone. Including morphological count data enables us to
328 estimate the confounded parameters from metabarcoding data. For each sampled station, we have
329 two independent sets of observed data: 1) counts of larval/juvenile fishes for each taxon from the
330 formaldehyde jars and 2) counts of amplicons for each taxon from ethanol jars. These observed
331 data arise from a common (but unobserved) biomass for each species at each station-year
332 combination, and thus we can model them jointly (71).

333 Microscopy counts were modeled as Poisson-distributed given their relatively small
334 absolute values and low variance (64), and amplicon sequence data were modeled using a
335 Negative Binomial distribution given their relatively high absolute values and high variability
336 among replicates (Figure S2-S4). These statistical distributions are commonly used in models of
337 amplicon and count data, respectively (35–37). A full description of model implementation and
338 assumptions are provided in Supplement 2.

339 *Data Analysis*

340 After model estimation, we calculated mean abundance estimates (larvae counts per
341 standardized volume towed) per species per station per year. To explore species-specific sea
342 surface temperature (SST) relationships, we fit a Bayesian generalized linear model using log
343 (abundance) as the response variable and SST ($^{\circ}\text{C}$) as a continuous predictor variable. Models
344 were implemented for each species using Stan as implemented in R (75, 76). We then

345 summarized the affinity between each species and SST by calculating a T-statistic based on each
346 species' estimated coefficients (mean slope/standard deviation). We further plotted the estimated
347 slope for each “species grouping” by habitat associations derived from previous CalCOFI
348 research (See Supplement 1 methods)(63). We summed total log (abundance) per habitat
349 association per station per year and fit a Bayesian generalized linear model using log (total
350 abundance) as the response variable and SST (°C) as a continuous predictor variable.

351 We repeated the above analyses using a Bayesian binomial model using presence as the
352 response variable and SST (°C) as a continuous predictor variable across the data set to explore
353 occurrence relationships with temperature and identify warm- and cool- associated taxa. We set a
354 threshold of presence/absence based on the model using a threshold of < 0.01 larvae per
355 standardized volume to be considered absent within a station.

356 We further explored species occurrence and abundance relationships with SST by fitting
357 the above Bayesian generalized linear models with station as a random effect (See Supplement 1
358 results).

359 We visualized anchovy and sardine abundance over time by calculating the mean log
360 (abundance) of each species per station per year. We then plotted the mean log (abundance) of
361 each of the four stations while error bars represent the 95% confidence intervals observed for a
362 given species at a given station in that year.

363 To evaluate the effect of the marine heatwave (MHW) on CCLME fishes we compared
364 estimated species abundances before the MHW (1996-2013), to both during and after the MHW
365 (2014-2019), at each station respectively. We first calculated the mean abundance for each
366 species at each station for each model run. We then subtracted the means for each model run to
367 evaluate changes in MHW abundance per species per station per model run. We then calculated a

368 95% CI of change in MHW abundance per species to identify which species were significantly
369 different before vs. during and after the MHW at each station.

370 All data and code to conduct analyses and generate all figures are available on GitHub
371 (https://github.com/zjgold/CalCOFI_eDNA) and associated Google Drive link
372 (https://drive.google.com/drive/folders/12cU9mY_CWoro-x6Hgh_pgv_66zZEzm1h?usp=sharing) [will be replaced with a Dryad repository upon
374 acceptance].

375

376 **Acknowledgments**

377 *Funding*

378 We acknowledge NSF GRFP and GRIP [DEG No. 2015204395], UCLA La Kretz Center
379 for Conservation Genomics, Howard Hughes Medical Institute, and UCLA Department of
380 Ecology and Evolutionary Biology for funding this research.

381 *Author Contributions*

- 382 • Conceptualization: ZG, DK, KDG, ART, LRT, PHB
- 383 • Performed Research: ZG, RPK, AOS, RG, DK, ART, PHB, KMP
- 384 • Funding Acquisition: ZG, PHB, ART, KDG, DK
- 385 • Data Curation: ZG, RPK, AOS, RG
- 386 • Formal Analysis: ZG, RPK, AOS, RG, ART
- 387 • Writing – Original Draft Preparation: ZG, RPK, AOS, ART, KDG, RG, KMP, LRT, DK,
- 388 PHB

389 *Competing Interests*

390 The authors declare no competing interests.

391 *Data and Materials Availability*

392 All data needed to evaluate the conclusions in the paper are present in the paper and/or
393 the Supplementary Materials. All data and code to conduct analyses and generate all figures are
394 available on GitHub (https://github.com/zjgold/CalCOFI_eDNA) and associated Google Drive
395 link (https://drive.google.com/drive/folders/12cU9mY_CWoro-x6Hgh_pgv_66zZEzm1h?usp=sharing) [will be replaced with a Dryad repository upon
396 acceptance].

398

399 **References**

- 400 1. E. C. J. Oliver, S. E. Perkins-Kirkpatrick, N. J. Holbrook, N. L. Bindoff, 9. Anthropogenic
401 and natural influences on record 2016 marine heat waves. *Bull. Am. Meteorol. Soc.* **99**,
402 S44–S48 (2018).
- 403 2. T. L. Frölicher, C. Laufkötter, Emerging risks from marine heat waves. *Nat. Commun.* **9**
404 (2018), pp. 1–4.
- 405 3. T. P. Hughes, J. T. Kerry, A. H. Baird, S. R. Connolly, A. Dietzel, C. M. Eakin, S. F.
406 Heron, A. S. Hoey, M. O. Hoogenboom, G. Liu, M. J. McWilliam, R. J. Pears, M. S.
407 Pratchett, W. J. Skirving, J. S. Stella, G. Torda, Global warming transforms coral reef
408 assemblages. *Nature*. **556**, 492–496 (2018).
- 409 4. L. Rogers-Bennett, C. A. Catton, Marine heat wave and multiple stressors tip bull kelp
410 forest to sea urchin barrens. *Sci. Rep.* **9**, 1–9 (2019).
- 411 5. J. M. Nielsen, L. A. Rogers, R. D. Brodeur, A. R. Thompson, T. D. Auth, A. L. Deary, J.
412 T. Duffy-Anderson, M. Galbraith, J. A. Koslow, R. I. Perry, Responses of ichthyoplankton
413 assemblages to the recent marine heatwave and previous climate fluctuations in several

- 414 Northeast Pacific marine ecosystems. *Glob. Chang. Biol.* **27**, 506–520 (2021).
- 415 6. W. W. L. Cheung, T. L. Frölicher, Marine heatwaves exacerbate climate change impacts
416 for fisheries in the northeast Pacific. *Sci. Rep.* **10**, 1–10 (2020).
- 417 7. M. L. Pinsky, R. L. Selden, Z. J. Kitchel, Climate-Driven Shifts in Marine Species
418 Ranges: Scaling from Organisms to Communities. *Ann. Rev. Mar. Sci.* **12**, 153–179
419 (2020).
- 420 8. E. A. Becker, K. A. Forney, J. V. Redfern, J. Barlow, M. G. Jacox, J. J. Roberts, D. M.
421 Palacios, Predicting cetacean abundance and distribution in a changing climate. *Divers.*
422 *Distrib.* **25**, 626–643 (2019).
- 423 9. D. M. Checkley, R. G. Asch, R. R. Rykaczewski, Climate, Anchovy, and Sardine. *Ann.*
424 *Rev. Mar. Sci.* **9**, 469–493 (2017).
- 425 10. F. P. Chavez, J. Ryan, S. E. Lluch-Cota, C. M. Niñuen, Climate: From anchovies to
426 sardines and back: Multidecadal change in the Pacific Ocean. *Science (80-.).* **299**, 217–
427 221 (2003).
- 428 11. M. Lindegren, D. M. Checkley, T. Rouyer, A. D. MacCall, N. C. Stenseth, Climate,
429 fishing, and fluctuations of sardine and anchovy in the California Current. *Proc. Natl.*
430 *Acad. Sci. U. S. A.* **110**, 13672–13677 (2013).
- 431 12. N. D. , E. Drenkard, A. R. Thompson, E. D. Weber, D. Wilson-Vandenberg, S.
432 McClatchie, J. A. Koslow, B. X. Semmens, Bridging From Monitoring to Solutions-Based
433 Thinking: Lessons From CalCOFI for Understanding and Adapting to Marine Climate
434 Change Impacts. *Front. Mar. Sci.* **6**, 695 (2019).
- 435 13. P. E. Smith, H. G. Moser, Long-term trends and variability in the larvae of Pacific sardine
436 and associated fish species of the California Current region. *Deep. Res. Part II Top. Stud.*

- 437 *Oceanogr.* **50**, 2519–2536 (2003).
- 438 14. C. H. Hsieh, C. S. Reiss, J. R. Hunter, J. R. Beddington, R. M. May, G. Sugihara, Fishing
439 elevates variability in the abundance of exploited species. *Nature*. **443**, 859–862 (2006).
- 440 15. A. R. Thompson, W. Watson, S. McClatchie, E. D. Weber, Multi-scale sampling to
441 evaluate assemblage dynamics in an oceanic marine reserve. *PLoS One*. **7**, e33131 (2012).
- 442 16. H. G. Moser P.E. Smith, and L.E. Eber, Larval fish assemblages in the California Current
443 region, 1954-1960, a period of dynamic environmental change. *CalCOFI Rep.* **28**, 97–127
444 (1987).
- 445 17. H. Moser, R. Charter, P. Smith, D. Ambrose, W. Watson, S. Charter, E. Sandknop,
446 Distributional atlas of fish larvae and eggs in the Southern California Bight region: 1951-
447 1998. *Calif. Coop. Ocean. Fish. Investig. Atlas*. **34**, 1–166 (2001).
- 448 18. M. A. Snyder, L. C. Sloan, N. S. Diffenbaugh, J. L. Bell, Future climate change and
449 upwelling in the California Current. *Geophys. Res. Lett.* **30**, 1823 (2003).
- 450 19. R. G. Asch, Climate change and decadal shifts in the phenology of larval fishes in the
451 California Current ecosystem. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E4065–E4074 (2015).
- 452 20. A. R. Thompson, N. J. Ben-Aderet, N. M. Bowlin, D. Kacev, R. Swalethorp, W. Watson,
453 Putting the Pacific marine heatwave into perspective: The response of larval fish off
454 southern California to unprecedented warming in 2014–2016 relative to the previous 65
455 years. *Glob. Chang. Biol.* **28**, 1766–1785 (2022).
- 456 21. A. R. Thompson, D. C. Chen, L. W. Guo, J. R. Hyde, W. Watson, Larval abundances of
457 rockfishes that were historically targeted by fishing increased over 16 years in association
458 with a large marine protected area. *R. Soc. Open Sci.* **4**, 170639 (2017).
- 459 22. Z. Gold, E. E. Curd, K. D. Goodwin, E. S. Choi, B. W. Frable, A. R. Thompson, H. J.

- 460 Walker, R. S. Burton, D. Kacev, L. D. Martz, P. H. Barber, Improving metabarcoding
461 taxonomic assignment: A case study of fishes in a large marine ecosystem. *Mol. Ecol.*
462 *Resour.* **21**, 2546–2564 (2021).
- 463 23. M. Miya, R. O. Gotoh, T. Sado, MiFish metabarcoding: a high-throughput approach for
464 simultaneous detection of multiple fish species from environmental DNA and other
465 samples. *Fish. Sci.* **86**, 939–970 (2020).
- 466 24. M. G. Jacox, M. A. Alexander, N. J. Mantua, J. D. Scott, G. Hervieux, R. S. Webb, F. E.
467 Werner, 6. Forcing of multiyear extreme ocean temperatures that impacted California
468 current living marine resources in 2016. *Bull. Am. Meteorol. Soc.* **99**, S27–S33 (2018).
- 469 25. C. A. Morgan, B. R. Beckman, L. A. Weitkamp, K. L. Fresh, Recent Ecosystem
470 Disturbance in the Northern California Current. *Fisheries.* **44**, 465–474 (2019).
- 471 26. C. L. Gentemann, M. R. Fewings, M. García-Reyes, Satellite sea surface temperatures
472 along the West Coast of the United States during the 2014–2016 northeast Pacific marine
473 heat wave. *Geophys. Res. Lett.* **44**, 312–319 (2017).
- 474 27. A. S. Ren, D. L. Rudnick, Temperature and salinity extremes from 2014–2019 in the
475 California Current System and its source waters. *Commun. Earth Environ.* **2**, 1–9 (2021).
- 476 28. I. D. Schroeder, J. A. Santora, S. J. Bograd, E. L. Hazen, K. M. Sakuma, A. M. Moore, C.
477 A. Edwards, B. K. Wells, J. C. Field, Source water variability as a driver of rockfish
478 recruitment in the California Current Ecosystem: implications for climate change and
479 fisheries management. *Can. J. Fish. Aquat. Sci.* **76**, 950–960 (2019).
- 480 29. M. Miya, Y. Sato, T. Fukunaga, T. Sado, J. Y. Poulsen, K. Sato, T. Minamoto, S.
481 Yamamoto, H. Yamanaka, H. Araki, M. Kondoh, W. Iwasaki, MiFish, a set of universal
482 PCR primers for metabarcoding environmental DNA from fishes: Detection of more than

- 483 230 subtropical marine species. *R. Soc. Open Sci.* **2**, 150088 (2015).
- 484 30. E. E. Curd, Z. Gold, G. S. Kandlikar, J. Gomer, M. Ogden, T. O'Connell, L. Pipes, T. M.
485 Schweizer, L. Rabichow, M. Lin, B. Shi, P. H. Barber, N. Kraft, R. Wayne, R. S. Meyer,
486 Anacapa Toolkit: An environmental DNA toolkit for processing multilocus metabarcode
487 datasets. *Methods Ecol. Evol.* **10**, 1469–1475 (2019).
- 488 31. R. Gallego, E. Jacobs-Palmer, K. Cribari, R. P. Kelly, Environmental DNA metabarcoding
489 reveals winners and losers of global change in coastal waters: EDNA and climate change.
490 *Proc. R. Soc. B Biol. Sci.* **287**, 20202424 (2020).
- 491 32. W. J. Sydeman, S. Dedman, M. García-Reyes, S. A. Thompson, J. A. Thayer, A. Bakun,
492 A. D. MacCall, Sixty-five years of northern anchovy population studies in the southern
493 California Current: A review and suggestion for sensible management. *ICES J. Mar. Sci.*
494 **77**, 486–499 (2020).
- 495 33. A. R. Thompson, I. D. Schroeder, S. J. Bograd, E. L. Hazen, M. G. Jacox, A. Leising, B.
496 K. Wells, J. L. Largier, J. L. Fisher, K. C. Jacobson, S. M. Zeman, E. P. Bjorkstedt, R. R.
497 Robertson, M. Kahru, R. Goericke, C. E. Peabody, T. Baumgartner, B. E. Lavaniegos, L.
498 E. Miranda, E. Gómez-Ocampo, J. Gómez-Valdés, T. D. Authy, E. A. Daly, C. A.
499 Morgan, J. B. Burke, J. C. Field, K. Sakuma, E. D. Weber, W. Watson, J. M. Porquez, J.
500 Dolliver, D. E. Lyons, R. A. Orben, J. Zamon, P. Warybok, J. Jahncke, J. A. Santora, S.
501 A. Thompson, B. Hoover, W. J. Sydeman, S. Melin, State of the California current 2018-
502 19 : a novel anchovy regime and a new marine heat wave? *Calif. Coop. Ocean. Fish.*
503 *Investig. Rep.* **60**, 1–60 (2019).
- 504 34. R. H. Parrish, D. L. Mallicoate, R. A. Klingbeil, Age dependent fecundity, number of
505 spawninge per year, sex ratio, and maturation stages in northern anchovy, *Engraulis*

- 506 mordax. *Fish. Bull.* **84**, 503–517 (1986).
- 507 35. E. D. Weber, T. D. Auth, S. Baumann-Pickering, T. R. Baumgartner, E. P. Bjorkstedt, S.
508 J. Bograd, B. J. Burke, J. L. Cadena-Ramírez, E. A. Daly, M. de la Cruz, State of the
509 California Current 2019–2020: Back to the Future With Marine Heatwaves? *Front. Mar.*
510 *Sci.*, 1081 (2021).
- 511 36. A. Vergés, C. Doropoulos, H. A. Malcolm, M. Skye, M. Garcia-Pizá, E. M. Marzinelli, A.
512 H. Campbell, E. Ballesteros, A. S. Hoey, A. Vila-Concejo, Y. M. Bozec, P. D. Steinberg,
513 Long-term empirical evidence of ocean warming leading to tropicalization of fish
514 communities, increased herbivory, and loss of kelp. *Proc. Natl. Acad. Sci. U. S. A.* **113**,
515 13791–13796 (2016).
- 516 37. C. Chaudhary, A. J. Richardson, D. S. Schoeman, M. J. Costello, Global warming is
517 causing a more pronounced dip in marine species richness around the equator. *Proc. Natl.*
518 *Acad. Sci. U. S. A.* **118** (2021), doi:10.1073/pnas.2015094118.
- 519 38. A. R. Thompson, C. J. Harvey, W. J. Sydeman, C. Barceló, S. J. Bograd, R. D. Brodeur, J.
520 Fiechter, J. C. Field, N. Garfield, T. P. Good, E. L. Hazen, M. E. Hunsicker, K. Jacobson,
521 M. G. Jacox, A. Leising, J. Lindsay, S. R. Melin, J. A. Santora, I. D. Schroeder, J. A.
522 Thayer, B. K. Wells, G. D. Williams, Indicators of pelagic forage community shifts in the
523 California Current Large Marine Ecosystem, 1998–2016. *Ecol. Indic.* **105**, 215–228
524 (2019).
- 525 39. J. A. Santora, N. J. Mantua, I. D. Schroeder, J. C. Field, E. L. Hazen, S. J. Bograd, W. J.
526 Sydeman, B. K. Wells, J. Calambokidis, L. Saez, D. Lawson, K. A. Forney, Habitat
527 compression and ecosystem shifts as potential links between marine heatwave and record
528 whale entanglements. *Nat. Commun.* **11**, 1–12 (2020).

- 529 40. H. Robinson, J. Thayer, W. J. Sydeman, M. Weise, Changes in California sea lion diet
530 during a period of substantial climate variability. *Mar. Biol.* **165**, 1–12 (2018).
- 531 41. J. F. Piatt, J. K. Parrish, H. M. Renner, S. K. Schoen, T. T. Jones, M. L. Arimitsu, K. J.
532 Kuletz, B. Bodenstein, M. García-Reyes, R. S. Duerr, R. M. Corcoran, R. S. A. Kaler, G.
533 J. McChesney, R. T. Golightly, H. A. Coletti, R. M. Suryan, H. K. Burgess, J. Lindsey, K.
534 Lindquist, P. M. Warzybok, J. Jahncke, J. Roletto, W. J. Sydeman, Extreme mortality and
535 reproductive failure of common murres resulting from the northeast Pacific marine
536 heatwave of 2014-2016. *PLoS One.* **15**, e0226087 (2020).
- 537 42. S. McClatchie, Sardine biomass is poorly correlated with the Pacific Decadal Oscillation
538 off California. *Geophys. Res. Lett.* **39** (2012), doi:10.1029/2012GL052140.
- 539 43. C. Deutsch, A. Ferrel, B. Seibel, H. O. Pörtner, R. B. Huey, Climate change tightens a
540 metabolic constraint on marine habitats. *Science (80-.)* **348**, 1132–1135 (2015).
- 541 44. K. A. Rose, J. Fiechter, E. N. Curchitser, K. Hedstrom, M. Bernal, S. Creekmore, A.
542 Haynie, S. ichi Ito, S. Lluch-Cota, B. A. Megrey, C. A. Edwards, D. Checkley, T. Koslow,
543 S. McClatchie, F. Werner, A. MacCall, V. Agostini, Demonstration of a fully-coupled
544 end-to-end model for small pelagic fish using sardine and anchovy in the California
545 Current. *Prog. Oceanogr.* **138**, 348–380 (2015).
- 546 45. J. A. Hare, The future of fisheries oceanography lies in the pursuit of multiple hypotheses.
547 *ICES J. Mar. Sci.* **71**, 2343–2356 (2014).
- 548 46. D. Robert, H. M. Murphy, G. P. Jenkins, L. Fortier, Poor taxonomical knowledge of larval
549 fish prey preference is impeding our ability to assess the existence of a “critical period”
550 driving year-class strength. *ICES J. Mar. Sci.* **71**, 2042–2052 (2014).
- 551 47. E. Garcia-Vazquez, O. Georges, S. Fernandez, A. Ardura, eDNA metabarcoding of small

- 552 plankton samples to detect fish larvae and their preys from Atlantic and Pacific waters.
- 553 *Sci. Rep.* **11**, 1–13 (2021).
- 554 48. K. J. Pitz, J. Guo, S. B. Johnson, T. L. Campbell, H. Zhang, R. C. Vrijenhoek, F. P.
555 Chavez, J. Geller, Zooplankton biogeographic boundaries in the California Current
556 System as determined from metabarcoding. *PLoS One.* **15**, e0235159 (2020).
- 557 49. C. Mariac, J. F. Renno, G. Carmen, Y. Vigouroux, E. Mejia, C. Angulo, D. Castro Ruiz,
558 G. Estivals, C. Nolorbe, A. García Vasquez, J. Nuñez, G. Cochonneau, M. Flores, J.
559 Alvarado, J. Vertiz, W. Chota-Macuyama, H. Sánchez, G. Miranda, F. Duponchelle,
560 Species-level ichthyoplankton dynamics for 97 fishes in two major river basins of the
561 Amazon using quantitative metabarcoding. *Mol. Ecol.* (2021), doi:10.1111/mec.15944.
- 562 50. M. Barbato, T. Kovacs, M. A. Coleman, M. K. Broadhurst, M. de Bruyn, Metabarcoding
563 for stomach-content analyses of Pygmy devil ray (*Mobula kuhlii* cf. *eregoodootenkee*):
564 Comparing tissue and ethanol preservative-derived DNA. *Ecol. Evol.* **9**, 2678–2687
565 (2019).
- 566 51. M. Erdozain, D. G. Thompson, T. M. Porter, K. A. Kidd, D. P. Kreutzweiser, P. K. Sibley,
567 T. Swystun, D. Chartrand, M. Hajibabaei, Metabarcoding of storage ethanol vs.
568 conventional morphometric identification in relation to the use of stream
569 macroinvertebrates as ecological indicators in forest management. *Ecol. Indic.* **101**, 173–
570 184 (2019).
- 571 52. H. L. Thalmann, E. A. Daly, R. D. Brodeur, Two Anomalously Warm Years in the
572 Northern California Current: Impacts on Early Marine Steelhead Diet Composition,
573 Morphology, and Potential Survival. *Trans. Am. Fish. Soc.* **149**, 369–382 (2020).
- 574 53. R. P. Kelly, A. O. Shelton, R. Gallego, Understanding PCR Processes to Draw

- 575 Meaningful Conclusions from Environmental DNA Studies. *Sci. Rep.* **9**, 1–14 (2019).
- 576 54. G. B. Gloor, J. M. Macklaim, V. Pawlowsky-Glahn, J. J. Egozcue, Microbiome datasets
577 are compositional: And this is not optional. *Front. Microbiol.* **8**, 2224 (2017).
- 578 55. M. R. McLaren, A. D. Willis, B. J. Callahan, Consistent and correctable bias in
579 metagenomic sequencing experiments. *Elife.* **8**, e46923 (2019).
- 580 56. J. D. Silverman, R. J. Bloom, S. Jiang, H. K. Durand, E. Dallow, S. Mukherjee, L. A.
581 David, Measuring and mitigating PCR bias in microbiota datasets. *PLoS Comput. Biol.* **17**,
582 e1009113 (2021).
- 583 57. M. Y. Stoeckle, J. Adolf, Z. Charlop-Powers, K. J. Dunton, G. Hinks, S. M. Vanmorter,
584 Trawl and eDNA assessment of marine fish diversity, seasonality, and relative abundance
585 in coastal New Jersey, USA. *ICES J. Mar. Sci.* **78**, 293–304 (2021).
- 586 58. A. Lacoursière-Roussel, G. Côté, V. Leclerc, L. Bernatchez, Quantifying relative fish
587 abundance with eDNA: a promising tool for fisheries management. *J. Appl. Ecol.* **53**,
588 1148–1157 (2016).
- 589 59. M. C. Yates, D. J. Fraser, A. M. Derry, Meta-analysis supports further refinement of
590 eDNA for monitoring aquatic species-specific abundance in nature. *Environ. DNA.* **1**, 5–
591 13 (2019).
- 592 60. M. M. Zenker, A. Specht, V. G. Fonseca, Assessing insect biodiversity with automatic
593 light traps in Brazil: Pearls and pitfalls of metabarcoding samples in preservative ethanol.
594 *Ecol. Evol.* **10**, 2352–2366 (2020).
- 595 61. S. McClatchie, A. R. Thompson, S. R. Alin, S. Siedlecki, W. Watson, S. J. Bograd, The
596 influence of Pacific Equatorial Water on fish diversity in the southern California Current
597 System. *J. Geophys. Res. Ocean.* **121**, 6121–6136 (2016).

- 598 62. H. G. Moser, R. L. Charter, P. E. Smith, D. A. Ambrose, S. R. Charter, C. A. Meyer, E.
599 M. Sandknop, W. Watson, *Distributional atlas of fish larvae and eggs in the California*
600 *Current region: taxa with 1000 or more total larvae, 1951 through 1984* (Marine Life
601 Research Program, Scripps Institution of Oceanography, 1993), vol. 53.
- 602 63. C. H. Hsieh, C. Reiss, W. Watson, M. J. Allen, J. R. Hunter, R. N. Lea, R. H. Rosenblatt,
603 P. E. Smith, G. Sugihara, A comparison of long-term trends and variability in populations
604 of larvae of exploited and unexploited fishes in the Southern California region: A
605 community approach. *Prog. Oceanogr.* **67**, 160–185 (2005).
- 606 64. A. R. Thompson, S. McClatchie, E. D. Weber, W. Watson, C. E. Lennert-Cody,
607 Correcting for bias in calcofi ichthyoplankton abundance estimates associated with the
608 1977 transition from ring to bongo net sampling. *Calif. Coop. Ocean. Fish. Investig.*
609 *Reports.* **58**, 1–11 (2017).
- 610 65. D. Kramer, M. J. Kalin, E. G. Stevens, J. R. Thrailkill, J. . Zweifel, *Collecting and*
611 *processing data on fish eggs and larvae in the California Current. NOAA Tech. Rep.*
612 *NMFS Circ.*, vol. 370. (US Department of Commerce, National Oceanic and Atmospheric
613 Administration ..., 1972), vol. 370.
- 614 66. S. McClatchie, *Regional fisheries oceanography of the California current system: The*
615 *CalCOFI program* (Springer, 2014).
- 616 67. M. Miya, Y. Sato, T. Fukunaga, T. Sado, J. Y. Poulsen, K. Sato, T. Minamoto, S.
617 Yamamoto, H. Yamanaka, H. Araki, M. Kondoh, W. Iwasaki, MiFish, a set of universal
618 PCR primers for metabarcoding environmental DNA from fishes: Detection of more than
619 230 subtropical marine species. *R. Soc. Open Sci.* **2**, 150088 (2015).
- 620 68. N. Lalam, Estimation of the reaction efficiency in polymerase chain reaction. *J. Theor.*

- 621 *Biol.* **242**, 947–953 (2006).
- 622 69. J. D. Silverman, K. Roche, S. Mukherjee, L. A. David, Naught all zeros in sequence count
623 data are the same. *Comput. Struct. Biotechnol. J.* **18**, 2789–2798 (2020).
- 624 70. J. J. Jos', J. Egozcue, J. Graffelman, M. I. Ortego, V. Pawlowsky-Glahn, Some thoughts
625 on counts in sequencing studies. *NAR Genomics Bioinforma.* **2**, 1–10 (2020).
- 626 71. N. T. Hobbs, M. B. Hooten, *Bayesian Models: A Statistical Primer for Ecologists.* 593
627 Princeton University Press (Princeton, 2015).
- 628 72. X. Ren, P. F. Kuan, Negative binomial additive model for RNA-Seq data analysis. *BMC*
629 *Bioinformatics.* **21**, 1–15 (2020).
- 630 73. T. Chambert, D. S. Pilliod, C. S. Goldberg, H. Doi, T. Takahara, An analytical framework
631 for estimating aquatic species density from environmental DNA. *Ecol. Evol.* **8**, 3468–3477
632 (2018).
- 633 74. E. Meyer-Gutbrod, L. Kui, R. Miller, M. Nishimoto, L. Snook, M. Love, Moving on up:
634 Vertical distribution shifts in rocky reef fish species during climate-driven decline in
635 dissolved oxygen from 1995 to 2009. *Glob. Chang. Biol.* **27**, 6280–6293 (2021).
- 636 75. B. Goodrich, J. Gabry, I. Ali, S. Brilleman, rstanarm: Bayesian applied regression
637 modeling via Stan. *R Packag. version.* **2**, 1758 (2020).
- 638 76. Stan Development Team, RStan: the R interface to Stan (2021).
- 639 77. J. N. Heine, California Cooperative Oceanic Fisheries Investigations. Reports. **49**, 264
640 (2008).
- 641 78. S. McClatchie, J. Gao, E. J. Drenkard, A. R. Thompson, W. Watson, L. Ciannelli, S. J.
642 Bograd, J. T. Thorson, Interannual and Secular Variability of Larvae of Mesopelagic and
643 Forage Fishes in the Southern California Current System. *J. Geophys. Res. Ocean.* **123**,

- 644 6277–6295 (2018).
- 645 79. E. Valsecchi, J. Bylemans, S. J. Goodman, R. Lombardi, I. Carr, L. Castellano, A.
- 646 Galimberti, P. Galli, Novel universal primers for metabarcoding environmental DNA
- 647 surveys of marine mammals and other marine vertebrates. *Environ. DNA.* **2**, 460–476
- 648 (2020).
- 649 80. R. A. Collins, G. Trauzzi, K. M. Maltby, T. I. Gibson, F. C. Ratcliffe, J. Hallam, S.
- 650 Rainbird, J. Maclaine, P. A. Henderson, D. W. Sims, S. Mariani, M. J. Genner, Meta-Fish-
- 651 Lib: A generalised, dynamic DNA reference library pipeline for metabarcoding of fishes.
- 652 *J. Fish Biol.* **99**, 1446–1454 (2021).
- 653 81. A. Polanco F., E. Richards, B. Flück, A. Valentini, F. Altermatt, S. Brosse, J. C. Walser,
- 654 D. Eme, V. Marques, S. Manel, C. Albouy, T. Dejean, L. Pellissier, Comparing the
- 655 performance of 12S mitochondrial primers for fish environmental DNA across
- 656 ecosystems. *Environ. DNA.* **3**, 1113–1127 (2021).
- 657 82. K. Deiner, H. M. Bik, E. Mächler, M. Seymour, A. Lacoursière-Roussel, F. Altermatt, S.
- 658 Creer, I. Bista, D. M. Lodge, N. de Vere, M. E. Pfrender, L. Bernatchez, Environmental
- 659 DNA metabarcoding: Transforming how we survey animal and plant communities. *Mol.*
- 660 *Ecol.* **26**, 5872–5895 (2017).
- 661 83. P. Taberlet, A. Bonin, L. Zinger, E. Coissac, *Environmental DNA: For biodiversity*
- 662 *research and monitoring* (Oxford University Press, 2018).
- 663 84. R. C. Edgar, Accuracy of taxonomy prediction for 16S rRNA and fungal ITS sequences.
- 664 *PeerJ.* **2018**, e4652 (2018).
- 665 85. M. A. Min, P. H. Barber, Z. Gold, MiSebastes: An eDNA metabarcoding primer set for
- 666 rockfishes (genus *Sebastodes*). *Conserv. Genet. Resour.* **13**, 447–456 (2021).

- 667 86. M. Leray, J. Y. Yang, C. P. Meyer, S. C. Mills, N. Agudelo, V. Ranwez, J. T. Boehm, R.
668 J. Machida, A new versatile primer set targeting a short fragment of the mitochondrial
669 COI region for metabarcoding metazoan diversity: Application for characterizing coral
670 reef fish gut contents. *Front. Zool.* **10**, 34 (2013).
- 671 87. P. A. Hastings, R. S. Burton, Establishing a DNA Sequence database for the marine fish
672 fauna of California, 5 (2008).
- 673 88. B. E. Deagle, S. N. Jarman, E. Coissac, F. Pompanon, P. Taberlet, DNA metabarcoding
674 and the cytochrome c oxidase subunit I marker: Not a perfect match. *Biol. Lett.* **10**,
675 20140562 (2014).
- 676 89. J. L. O'donnell, R. P. Kelly, N. C. Lowell, J. A. Port, Indexed PCR primers induce
677 template- Specific bias in Large-Scale DNA sequencing studies. *PLoS One.* **11**, e0148698
678 (2016).
- 679 90. D. M. Gohl, P. Vangay, J. Garbe, A. MacLean, A. Hauge, A. Becker, T. J. Gould, J. B.
680 Clayton, T. J. Johnson, R. Hunter, D. Knights, K. B. Beckman, Systematic improvement
681 of amplicon marker gene methods for increased accuracy in microbiome studies. *Nat.*
682 *Biotechnol.* **34**, 942–949 (2016).
- 683 91. K. Bohmann, V. Elbrecht, C. Carøe, I. Bista, F. Leese, M. Bunce, D. W. Yu, M. Seymour,
684 A. J. Dumbrell, S. Creer, Strategies for sample labelling and library preparation in DNA
685 metabarcoding studies. *Mol. Ecol. Resour.* (2021), doi:10.1111/1755-0998.13512.
- 686 92. R. P. Kelly, R. Gallego, E. Jacobs-Palme, The effect of tides on nearshore environmental
687 DNA. *PeerJ.* **2018**, e4521 (2018).
- 688 93. R. Mendelssohn, rerddapXtracto: Extracts Environmental Data from “ERDDAP” Web
689 Services. R package version 1.0.0 (2020).

- 690 94. H. G. Moser, R. L. Charter, W. Watson, D. A. Ambrose, K. T. Hill, P. E. Smith, J. L.
691 Butler, E. M. Sandknop, S. R. Charter, The CalCOFI ichthyoplankton time series:
692 Potential contributions to the management of rocky-shore fishes. *Calif. Coop. Ocean.*
693 *Fish. Investig. Reports.* **42**, 112–128 (2001).
- 694 95. S. Juggins, rioja: Analysis of Quaternary science data, R package version (0.9-9). *Compr.*
695 *r Arch. Netw.* (2015).
- 696 96. A. de Vries, B. D. Ripley, Create Dendograms and Tree Diagrams Using “ggplot2.” *URL*
697 <https://github.com/andrie/ggdendro>, 12 (2020).
- 698 97. J. Oksanen, F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, O. R.B., G. L. Simpson,
699 P. Solymos, M. H. H. Stevens, H. Wagner, *vegan*: Community ecology package. *R*
700 *Packag. version 2.3-5* (2016), (available at <https://cran.r-project.org/package=vegan>).
- 701 98. J. Niku, F. K. C. Hui, S. Taskinen, D. I. Warton, gllvm: Fast analysis of multivariate
702 abundance data with generalized linear latent variable models in r. *Methods Ecol. Evol.*
703 **10**, 2173–2182 (2019).
- 704 99. A. Kucukelbir, R. Ranganath, A. Gelman, D. Blei, Automatic variational inference in
705 *Stan. Adv. Neural Inf. Process. Syst.* **28** (2015).
- 706 100. R. Jiang, W. V. Li, J. J. Li, mbImpute: an accurate and robust imputation method for
707 microbiome data. *Genome Biol.* **22**, 1–27 (2021).
- 708 101. M. Leray, N. Knowlton, Random sampling causes the low reproducibility of rare
709 eukaryotic OTUs in Illumina COI metabarcoding. *PeerJ.* **2017**, e3006 (2017).
- 710 102. J. A. Royle, W. A. Link, Generalized site occupancy models allowing for false positive
711 and false negative errors. *Ecology.* **87**, 835–841 (2006).
- 712

1 **SUPPLEMENT 1**

2 **Table of Contents**

3 **1. Methods**

- 4 a. *Study Design*
5 b. *Metabarcoding Collection Isolation, Amplification, and Sequencing*
6 c. *Bioinformatics*
7 d. *Microscopy Identification of Ichthyoplankton*
8 e. *Estimating Abundance (see Supplement 2)*
9 f. *Environmental Covariates*
10 g. *Data analysis*

11 **2. Results**

- 12 a. *Change in Fish Assemblage Structure over Time*
13 b. *Metabarcoding Signal Does Not Degrade Over Time*

14 **3. Figures**

- 15 a. *Figure S1. station Map*
16 b. *Figure S2. Observed Sequence Reads vs. Observed Morphological Counts*
17 c. *Figure S3. Predicted Counts vs. Observed Morphological Counts*
18 d. *Figure S4. Predicted Sequence Reads vs. Observed Sequence Reads*
19 e. *Figure S5. Co-detection of Taxa By Metabarcoding and Microscopy*
20 f. *Figure S6. Temperature Associations in Fish Species*
21 g. *Figure S7. Southern Oceanic Species Drive Fish Community Shifts*
22 h. *Figure S8. Significant Species Occurrence and SST Correlations at each Station*

- 23 i. *Figure S9. Significant Species Occurrence and MWCT Correlations at each*
24 *Station*
- 25 j. *Figure S10. Significant Species Abundance and SST Correlations at each station*
- 26 k. *Figure S11. Significant Species Abundance and MWCT Correlations at each*
27 *station*
- 28 l. *Figure S12. Bar Plot of Significant Species Abundance and SST Correlations*
29 *Across All stations*
- 30 m. *Figure S13. Bar Plot of Significant Species Abundance and MWCT Correlations*
31 *Across All stations*
- 32 n. *Figure S14. Increased Abundance of Southern Mesopelagic Species with Higher*
33 *SST*
- 34 o. *Figure S15. Increased Abundance of Southern and Central Mesopelagic Species*
35 *with Higher SST*
- 36 p. *Figure S16. Heat Map of Abundances Over Time*
- 37 q. *Figure S17. NMDS Ordination of Species and Years*
- 38 r. *Figure S18. NMDS Ordination of Species and Samples*
- 39 s. *Figure S19. Heat Map of San Diego Offshore Abundances Over Time*
- 40 t. *Figure S20. NMDS Ordination of San Diego Offshore Species and Years*
- 41 u. *Figure S21. Heat Map of San Diego Inshore Abundances Over Time*
- 42 v. *Figure S22. NMDS Ordination of San Diego Inshore Species and Years*
- 43 w. *Figure S23. Heat Map of Pt. Conception Abundances Over Time*
- 44 x. *Figure S24. NMDS Ordination of Pt. Conception Species and Years*
- 45 y. *Figure S25. Heat Map of San Nicholas Island Abundances Over Time*

46 z. *Figure S26. NMDS Ordination of San Nicholas Island Species and Years*
47 aa. *Figure S27. Co-occurrence Patterns of Species Controlling for SST*
48 bb. *Figure S28. Co-occurrence Patterns of Species*
49 cc. *Figure S39. Differential Species Abundance Before and After the Marine Heat*
50 *Wave at the San Nicholas Island*
51 dd. *Figure S30. Stable Precision of Amplicon Abundance Over Time*
52 ee. *Figure S31. Stable Precision of Abundance Estimates Over Time*

53

54 **Introduction**

55 This supplemental material provides additional details on the methods, results, and
56 discussion to support the main findings and conclusion of the manuscript.

57 The CalCOFI program (<https://calcofi.com/>) serves to provide fisheries-independent
58 ecosystem assessments of fish assemblages in the Southern California Current and has provided
59 decades of data on ichthyoplankton assemblages (77). The current CalCOFI surveys sample four
60 times per year from the U.S. Mexican Border to Monterey Bay (77). We used this rich sample
61 archive to interrogate ichthyoplankton assemblages from 1996-2019 (See Supplemental
62 Methods). Importantly, the occurrence of a marine heatwave (MHW) within the study region and
63 sampling period provided an additional opportunity to investigate the utility of having a non-
64 destructive means of interrogating the valuable CalCOFI sample archive.

65

66 **Methods**

67 *Study Design*

68 To evaluate the efficacy of metabarcoding methods used to analyze ethanol preserved
69 samples and investigate potential changes in the ichthyoplankton assemblages over decadal
70 scales, we identified ichthyoplankton by metabarcoding and microscopy in ethanol-preserved
71 samples collected over two decades (1996, 1998-2019; Figure S5) of spring CalCOFI cruises (5,
72 27, 35, 38). We note that samples collected in 1997 were stored in <50% ethanol and were
73 discarded due to failed preservation.

74 Samples were collected in late March or early April of each year
75 (calcofi_metadata_analysis_20210907.csv). Here we focus on spring samples because the
76 majority of species in the California Current spawn in spring and historically the annual
77 California Current Ecosystem Report has relied on the spring data (20, 78). This decision is
78 supported by recent work using ichthyoplankton data across the full set of yearly CalCOFI
79 cruises which found little evidence for phenological trends (20, 78), thus aiding the ability to
80 look at impacts from the marine heatwave.

81 Samples were collected from four well-separated stations (up to 370 km apart) from
82 distinct vicinities of the California Current with differing water properties (61) (5, 27, 35, 38)
83 (Figure S1). The northernmost station was located offshore of Point Conception, CA within the
84 California Current (34.14833°N -121.1567°W). The second station was located off San Nicholas
85 Island, CA (33.32333 °N, -119.6667°W) that experiences high variation in annual temperature
86 depending on the respective strengths of the California Current and Southern California Counter
87 Current (61). The third station was a southern coastal inshore station off San Diego, CA
88 (32.84667°N, -117.5383°W) characterized by relatively warmer waters from the California
89 Counter Current with seasonal (spring) upwelling of cool, nutrient-rich water (61). The fourth

90 station was a southern offshore station (31.85000°N , $-119.5683^{\circ}\text{W}$) characterized by sub-tropical
91 oceanic waters (Figure 4).

92 At each station, oblique bongo net tows were conducted from 210 m to the surface using
93 standard CalCOFI methods (15, 64–66). Each side of the bongo net had a 0.71 m-diameter
94 mouth opening and a net size of 0.505 mm mesh. Cod end contents of both bongo nets were
95 preserved at sea. The starboard side was preserved in sodium borate-buffered 2% formaldehyde
96 and the port side was preserved in Tris-buffered 95% ethanol. Ethanol was replaced after 24
97 hours to account for dilution from tissue water loss. Microscopy was conducted to identify
98 species abundance from formaldehyde-preserved samples following standardized CalCOFI
99 techniques (61) while metabarcoding was conducted on the ethanol in which port side samples
100 were stored; consequently, we expected the contents of the paired samples to differ slightly as a
101 function of sampling stochasticity.

102 *Metabarcoding Collection Isolation, Amplification, and Sequencing*

103 Prior to filtration, the ethanol-preserved samples were inverted three times and let rest for
104 30 minutes to resuspend and homogenize samples in the preservative. Filtration of ethanol from
105 the port-side bongo samples was conducted in a pre-PCR clean room at the NOAA Southwest
106 Fisheries Science Center within a biological safety cabinet in July 2019. The pre-PCR room had
107 no previous post-PCR work conducted within and all surfaces and equipment were sterilized
108 frequently with 10% bleach and 70% ethanol. The pre-PCR clean room was at ambient pressure
109 and reasonable precautions to limit contamination were conducted including only wearing clean
110 clothes that have not been exposed to labs with PCR product, no food brought into the lab, and
111 gloves were exchanged regularly.

112 Ethanol preservative was filtered using a vacuum filtration manifold with Nalgene
113 Analytical Test Filter Funnels (Thermofisher Scientific, Waltham, MA, USA) with the
114 manufacturer's 0.45 µm filters replaced with 0.2 µm Durapore PVDF filters (Sigma Aldrich, St.
115 Louis, MO, USA) using sterile forceps. Up to 125 mL of ethanol was then transferred from the
116 preserved jars into the filter funnels using a 10 mL pipette, carefully avoiding any sample
117 contents and thus preserving CalCOFI specimens for future research and analysis. Sample jars
118 were refilled using freshly prepared tris-buffered ethanol before being returned to the collection
119 archive. We included two negative controls to test for lab contamination by filtering 125 mL of
120 molecular grade water. Filters were stored at -20°C before DNA extraction.

121 Filters were extracted using the standard Qiagen DNAeasy Kit (Qiagen Inc., Valencia,
122 CA, USA) in a pre-PCR molecular lab. Extracted DNA was amplified using the MiFish
123 Universal Teleost primer sets to capture fish diversity (67).

124 Here, we highlight our decision to utilize the MiFish Universal Teleost *12S* primers. First,
125 these primers have been rigorously validated for fish barcoding (22, 23, 29, 30, 79–81) and
126 shown to provide accurate taxonomic assignments for a broad range of fishes (22). We recognize
127 that there are limitations for this, and indeed all, metabarcoding primer sets (82) which are forced
128 to balance specificity [how well target species can be taxonomically resolved] against breadth
129 [range of species across the tree of life that can be amplified] (83). Even a “gold standard” like
130 the *16S* rRNA gene marker for prokaryotic sequences struggles with taxonomic assignment
131 accuracy (84), especially with short-read sequences. Although taxonomic resolution limitations
132 and compromises remain for the *12S* target (22, 85), the taxonomic resolution has been improved
133 and best practices for taxonomic classification have been identified through the development of a

134 nearly comprehensive California Current Large Marine Ecosystem *12S* reference database along
135 with a full factorial cross-validation analysis of bioinformatic approaches (22).

136 Second, there are no widely used or benchmarked *COI* metabarcoding primer sets for
137 fish applications although *COI* barcoding is a common barcoding target. This is because a) the
138 conserved nature of the locus across the tree of life which results in amplification of a broad
139 array of taxa (86, 87), and b) the mismatch in high throughput sequencing platform length (max
140 is paired-end 300 bp) and rate of *COI* evolution/accumulation of sequence differences between
141 species (81, 88). In fact, these shortcomings were the original motivation for researchers to
142 develop alternative fish metabarcoding loci targeting *12S* loci for fishes (23). Together, the
143 research community has largely converged on the MiFish Universal Teleost *12S* primer set as
144 standard practice for fish metabarcoding given its balance of high specificity and breadth (23).
145 Thus we feel confident that the MiFish Universal Teleost *12S* primer set was an appropriate
146 choice for metabarcoding here.

147 Each metabarcoding extraction was subsampled for three PCR reactions using the MiFish
148 *12S* primer set. PCR amplification for the MiFish primer set was conducted following the
149 thermocycler profile of Curd *et al.* (30). MiFish PCR reactions had 25 µL reaction volume
150 containing 12.5 µL QIAGEN Multiplex Taq PCR 2x Master Mix (Qiagen Inc., Valencia, CA,
151 USA), 6.5 µL of molecular grade water, 2.5 µL of each primer (2 µmol/L), and 1 µL DNA
152 extraction. MiFish PCR thermocycling employed a touchdown profile with an initial
153 denaturation at 95°C for 15 min to activate the DNA polymerase, followed by 13 cycles of a 30s
154 denaturation at 94°C, a 30s annealing that started at 69.5°C and then decreased by 1.5°C for each
155 subsequent cycle (last cycle was 50°C), finishing with a 1 min extension at 72°C. This initial

156 touchdown profile was followed by 35 additional cycles using identical parameters except a
157 constant annealing temperature of 50°C and ending with a final extension at 72°C for 10 min.

158 Two non-native non-marine vertebrates, American alligator (*Alligator mississippiensis*)
159 and dromedary camel (*Camelus dromedarius*), were purchased at a local market and used as
160 positive controls. For all positive controls, tissues were extracted using the Qiagen Blood and
161 Tissue kit following the manufacturer's instructions. All PCR products were visualized via
162 electrophoresis on 2% agarose gels to ensure amplification success and correct product size.
163 Only filters from four jars failed to amplify, and upon further inspection within the archived
164 notes, all these samples had known preservation issues (e.g., preservative dried out, observed
165 mold, etc.). All other DNA extractions successfully amplified.

166 We prepared libraries following the methods of Curd *et al.* using a two-step PCR
167 amplification method with one final pool per primer set. Previous work indicated that two-step
168 PCR amplification can reduce amplification biases (89, 90) perhaps introduced by the inclusion
169 of various indices during one-step PCR procedures. Variations in the relative amplification
170 efficiency of each PCR is a concern here given the desire to study an array of targets in an
171 oceanic region over space and time. Overall, there are review papers available that outline the
172 advantages and disadvantages for one-step and two-step PCR protocols (91).

173 Prior to the second indexing PCR reaction, PCR samples from the first reaction were
174 cleaned using the Serapure magnetic bead protocol. We quantified bead-cleaned samples with
175 the Quant-iT™ broad range dsDNA Assay Kit (Thermofisher Scientific, Waltham, MA, USA)
176 on a Victor3 plate reader (Perkin Elmer Waltham, MA, USA). We indexed the sample libraries
177 using unique combinations of the Nextera Index A, B, C, and D Kit (Illumina, San Diego, CA,
178 USA) and KAPA HiFi HotStart Ready Mix (Kapa Biosystems, Sigma Aldrich, St. Louis, MO,

179 USA). Indexing was performed with a second PCR using a 25 µL reaction mixture containing
180 12.5 µL of Kapa HiFi Hotstart Ready mix, 1.25 µL of index primers, 10 ng of template DNA to
181 ensure equal copy number, and the remaining volume was filled using molecular grade water
182 depending on cleaned PCR product concentration. Index thermocycling parameters were:
183 denaturation at 95°C for 5 min, 5 cycles of denaturation at 98°C for 20 sec, annealing at 56°C for
184 30 sec, extension at 72°C for 3 min, followed by a final extension at 72°C for 5 min. To confirm
185 successful PCR and correct product size, we electrophoresed PCR products on 2% agarose gels.
186 We then bead cleaned and quantified DNA concentration, as described above so that we could
187 pool samples so as to have equal copy number for each unique library. Pooled libraries were
188 sequenced on an Illumina NextSeq PE 2x150 at UCLA Technology Center for Genomics and
189 Bioinformatics.

190 *Bioinformatics*

191 The resulting metabarcoding data were processed using the *Anacapa Toolkit* to conduct
192 quality control, amplicon sequence variant (ASV) parsing, and taxonomic assignment using user-
193 generated custom reference databases. We processed sequences using default parameters except
194 using a Q score cutoff of 30 and assigned taxonomy using *CRUX*-generated metabarcode
195 specific reference databases (22). The MiFish sequencing data was assigned taxonomy using the
196 California fish specific reference database and a bootstrap confidence cutoff score of 60
197 following Gold *et al.* (22).

198 The two resulting raw ASV community tables were decontaminated following Kelly *et al.*
199 (92). First, only merged paired reads that occurred at least twice (e.g., no singletons) were
200 retained. Second, we estimated index hopping between samples by calculating the proportion of
201 sequences within the positive control samples and then subtracting reads from each sample by

202 the sample read depth multiplied by the proportion of reads observed in the positive controls.
203 Third, we discarded technical replicates with fewer than 30,000 reads. Fourth, we calculated
204 Bray-Curtis dissimilarities between technical PCR replicates and fit a skewed beta distribution
205 ($a= 0.6$, $b= 9.5$). We then removed all replicates with greater than 95% probability of belonging
206 to the beta distribution. Resulting tables were then combined into a final ASV community table
207 in *R*.

208 *Microscopy Identification of Ichthyoplankton*

209 Plankton samples were processed at the NOAA Southwest Fisheries Science Center
210 ichthyoplankton laboratory. From each plankton sample, fish larvae were sorted and identified
211 through microscopy to the lowest practical taxon (64, 66). Most taxa were identified to species
212 although some were only characterized to genus or family level (See
213 `larval_counts_20210305.csv`). The number of larvae per species per jar, total abundance of
214 filtered ichthyoplankton, and proportion of jar sorted were recorded.

215 *Estimating Abundance*

216 We estimated the abundance of ichthyoplankton in each jar using a novel joint Bayesian
217 hierarchical model described in detail in Supplement 2.

218 *Environmental Covariates*

219 We specifically examined the relationship of ichthyoplankton communities to sea surface
220 temperatures (SST). Two month prior mean SSTs were obtained using the *rerddapXtracto*
221 package (93) in *R* to collect PathFinder Ver 5.3 monthly remotely sensed composites. To
222 calculate two-month prior means we first obtained monthly composites from April 1995 to April
223 2019 for each station. We then averaged across monthly composite sea surface temperatures

224 ignoring any missing values. Prior two month sea surface temperatures were chosen given the
225 average age of spring larvae (94) (Figure 4).

226 We then investigated the relationship of ichthyoplankton assemblages against sea surface
227 temperature as a proxy for a multitude of environmental shifts associated with the MHW. In
228 addition, we specifically characterized fish that are uniquely present or absent before (1996-
229 2013) and after the 2014-16 Marine Heatwave (2014-2019). Analyses were repeated using the
230 mean average water column temperature (MWCT) obtained from nearly simultaneously
231 conducted CTD rosette deployments (15). The mean water column temperature was averaged
232 across 10 to 100m depth, where the majority of ichthyoplankton reside, following Thompson et
233 al. (15).

234 *Data Analysis*

235 After model estimation, we calculated mean abundance estimates (larvae counts per
236 standardized volume towed) per species per station per year. To explore species-specific sea
237 surface temperature (SST) relationships, we fit a Bayesian generalized linear model using log
238 (abundance) as the response variable and SST ($^{\circ}\text{C}$) as a continuous predictor variable. Models
239 were implemented for each species using Stan as implemented in R (75). We then summarized
240 the affinity between each species and SST by calculating a T-statistic based on each species'
241 estimated coefficients (mean slope/standard deviation). We further plotted the estimated slope
242 for each “species grouping” by habitat associations derived from previous CalCOFI research
243 (habitat_association_to_check_art.csv)(63). We summed total log(abundance) per habitat
244 association per station per year and fit a Bayesian generalized linear model using log (total
245 abundance) as the response variable and SST ($^{\circ}\text{C}$) as a continuous predictor variable.

246 We repeated the above analyses using a Bayesian binomial model using presence as the
247 response variable and SST ($^{\circ}\text{C}$) as a continuous predictor variable across the data set to explore
248 occurrence relationships with temperature and identify warm- and cool- associated taxa. We set a
249 threshold of presence/absence based on the model using a threshold of < 0.01 larvae per
250 standardized volume to be considered absent within a station.

251 We further explored species occurrence and abundance relationships with SST by fitting
252 the above Bayesian generalized linear models with station as a random effect. In addition, we
253 repeated all of the above analyses using MWCT instead of SST.

254 To explore how fish assemblages change over time we plotted a heatmap of observed
255 abundance summed across stations each year. Chronological clustering was conducted across
256 years using Bray Curtis dissimilarities of abundances using a K of 8 using the package *rioja* in R
257 (95) and a dendrogram of years was constructed using the *ggdenro* package (96). Similarly,
258 hierarchical clustering was conducted across species using Bray Curtis dissimilarities of
259 abundances using a K of 6. To further explore fish assemblage changes NMDS Ordination of
260 Bray-Curtis dissimilarities were calculated from estimated abundances of each year summed
261 across stations as implemented by the *metaMDS* function from *vegan* in R (97). The above
262 analyses were also conducted with station separated as well as each station on it's own. To
263 investigate the relative effect of year, SST, and station to the explained variance in fish
264 assemblages across the ata set, we ran a PERMANOVA on Bray-Curtis dissimilarities using the
265 following model: ~ Year + SST + station.

266 We visualized anchovy and sardine abundance over time by calculating the median log
267 (abundance) of each species per station per year. We then plotted the log (median) abundance of

268 each of the four stations while error bars represent the 95% confidence intervals observed for a
269 given species at a given station in that year.

270 To explore co-occurrence patterns across species, we fit a generalized linear latent
271 variable model (GLLVM) following the methods of Niku et al. (98), specifically conducting
272 model fitting to determine the best distribution fit as well as number of latent variables to use.

273 The highest performing GLLVM employed 3 latent variables and applied a negative binomial
274 distribution with variational approximation (99) on the joint model predicted larvae count data.

275 We then plotted the correlation matrix of the linear predictors across species with and without
276 incorporating SST in the GLLVM to identify co-occurring species and the effect of SST on co-
277 occurrence patterns.

278 To evaluate the effect of the marine heatwave (MHW) on CCLME fishes we compared
279 estimated species abundances before the MHW (1996-2013), to both during and after the MHW
280 (2014-2019), at each station respectively. We first calculated the mean abundance for each
281 species at each station for each model run. We then subtracted the means for each model run to
282 evaluate changes in MHW abundance per species per station per model run. We then calculated a
283 95% CI of change in MHW abundance per species to identify which species were significantly
284 different before vs. during and after the MHW at each station.

285 All data and code to conduct analyses and generate all figures are available on GitHub
286 (https://github.com/zjgold/CalCOFI_eDNA) and associated Google Drive link
287 (https://drive.google.com/drive/folders/12cU9mY_CWoro-x6Hgh_pgv_66zZEzm1h?usp=sharing) [will be replaced with a Dryad repository upon
289 acceptance].

290

291 **Results**

292 *Fish Assemblage Structure*

293 We observed substantial changes in fish assemblage structure across stations, time, and
294 temperature sampled (NMDS stress =0.03) (Figure S11-26). station explained the greatest
295 observed variance (12%) which is unsurprising given the intentionally chosen distinct
296 biogeographic characteristics of each station (PERMANOVA p <0.05). However, despite the >
297 370km distance between stations, we captured significant synchronous changes in fish
298 assemblage dynamics in response to year (2.4%) and temperature (4.6%) (PERMANOVA p
299 <0.05). In particular, we observed strong clustering of the post MHW period from 2017-2019,
300 the 2005 El Niño and the 1998 El Niño along with southern mesopelagic species. Both 2014 and
301 2016 were distinct from other years and associated with a suite of mesopelagic species, although
302 the MHW itself was not strongly clustered largely due to the differential onset and
303 characterization of the warming event within the region (5).

304 We also found strong positive and negative co-occurrence patterns through GLLVM
305 analyses across species when controlling for temperature. Specifically, we observed strong
306 negative co-occurrence patterns between fisheries targets like benthic fisheries targets (e.g.
307 *Citharichthys* sp. sanddabs) and mesopelagic fishes (S27). We also observed strong positive co-
308 occurrence patterns between a suite of benthic species as well as strong positive co-occurrence
309 patterns between a suite of mesopelagic fishes. These results suggest that when controlling for
310 temperature, we observe strong benthic versus pelagic tradeoffs as observed previously (13, 20,
311 62).

312 When focusing on co-occurrence patterns without controlling for temperature, we
313 observed strong negative co-occurrence patterns with fisheries targets (North Pacific Hake

314 *Merluccius productus*) and mesopelagic fishes (S28). Here, temperature explained 19% of the
315 variability among species co-occurrence and was particularly important in driving negative co-
316 occurrence patterns between North Pacific Hake and mesopelagic species. These results suggest
317 that temperature may mediate tradeoffs between fisheries versus southern mesopelagic fish
318 assemblages. Further work exploring the underlying mechanisms of these negative co-
319 occurrence patterns is warranted.

320 *Analysis Using Mean Water Column Temperature*

321 We repeated the analyses presented in the main manuscript with mean water column
322 temperature (MWCT) as opposed to two-month SST (Figures 1-3). Two-month averaged SST
323 and instantaneous MWCT were only 29% correlated (linear regression, $p < 0.01$). This finding is
324 perhaps unsurprising given the substantial difference in spatial and temporal integration time of
325 these temperature measurements. Despite these apparent differences, we found nearly identical
326 results in the direction and significance of species–temperature associations as well as
327 temperature-driven variation in fish assemblage dynamics (Figures S6–S13). These results
328 suggest that species-temperature associations and our conclusions are largely robust to
329 temperature metrics.

330 *Metabarcoding Signal Appears Stable in the Ethanol-Preserved Samples*

331 For each station-species combination, if metabarcoding signals appear auto-correlated in
332 time -- that is, if one year's metabarcoding signal is correlated with the previous year's signal --
333 then we require a time-series model that incorporates such autocorrelation into the error
334 structure. If, by contrast, years appear independent of one another, we can treat model variation
335 as time-independent and therefore treat each data point as being independent. We observe no

336 such correlation (mean = -0.014, standard deviation = 0.35) and so we treat all observations as
337 independent of one another.

338 In further investigating the question of whether these samples can be considered time-
339 independent, we considered whether or not older samples might have less metabarcoding signal
340 due to sample degradation. If the metabarcoding signal were degrading overtime in the preserved
341 samples, we would expect several parameters to change as a function of sample age: (1) a
342 decrease in precision with which we observe amplicon abundance, (2) a decrease in richness of
343 species detected, and (3) a decrease in the confidence in posterior estimates of larval abundances
344 from our joint Bayesian model. We test for these effects in turn.

345 First, among triplicate PCR reactions, we might expect degraded DNA to behave more
346 stochastically than non-degraded DNA, such that technical replicates would yield increasingly
347 divergent amplicon abundances with greater degradation. Here, we measure the precision of our
348 estimates with the coefficient of variation (CV) of species-specific amplicons across three
349 technical replicates. An increase in CV with the age of the sample would signal degradation, but
350 we saw no such trend (Figure S29). Second, rare amplicons often make up a large fraction of
351 metabarcoding datasets, and because of their rarity, these often show up stochastically across
352 replicates or sequenced samples. If older DNA samples were degraded, we would expect fewer
353 of these rare species, and by extension, fewer species overall. We saw no such effect (linear
354 regression $p > 0.5$; linear mixed effect model failed convergence). Third, we might expect -- if
355 DNA were degrading -- that such degradation would impair our ability to estimate the larval
356 abundance of each species in older samples. Again, we saw no evidence of this effect (Figure
357 S30-31).

358 *Overlap in Species Detections*

359 The maximum observed morphological counts in which metabarcoding failed to detect a given
360 taxa was 9 (mean = 1.61). Across a total of 4,704 possible detections, 70.2% were non-detections
361 by both methods, 11.2% were detections by both methods, 16.4% were detections only made by
362 metabarcoding, and 2.1% were detections only made my microscopy (Figure S4).

363 *Stochasticity in Metabarcoding Data*

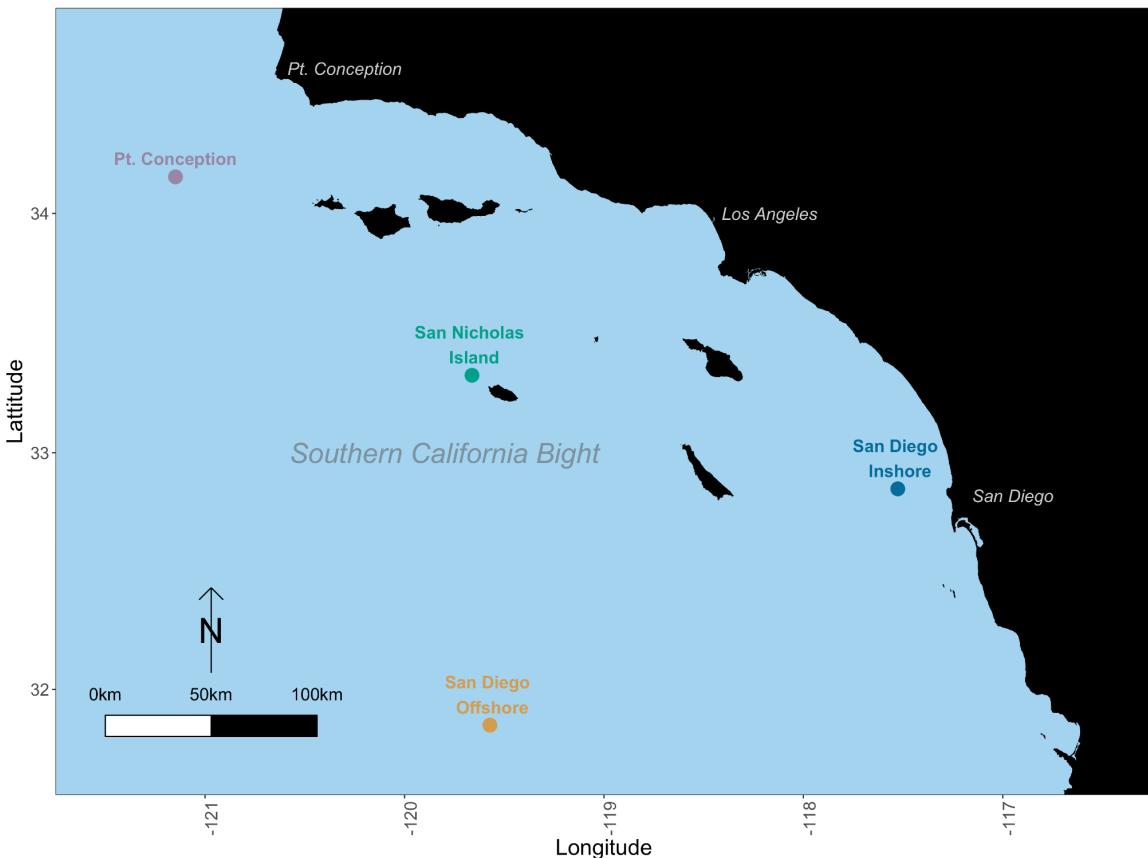
364 We conducted a deep dive into the origin and source of variation in amplicon sequence data.
365 These analyses identified stochastic dropouts in whereby a taxon is amplified in one PCR
366 reaction but not in a replicate PCR reaction as a main driver of variation in this data set. This is a
367 well-known phenomenon general to PCR with rare templates (69, 100, 101) and we document
368 such behavior in this manuscript (Figures S2-S4). For example, for *Symbolophorus*
369 *californiensis* we observed an instance of 3,897 reads, 165 reads, and 0 reads across three
370 technical PCR replicates with sample read depths of 132,731, 196,260, 55,400 from the same
371 DNA extraction. These stochastic dropouts are easily visualized along the X axis in Figure S2.
372 We note that the highest observed species-specific amplicon sample read proportion associated
373 with a stochastic dropout was 2.9% (3,897 /132,731) with the vast majority of such dropouts
374 occurring below 0.03% read proportion within a technical replicate. These results suggest that
375 stochasticity is largely driven by the abundance of DNA molecules within a sample rather than a
376 specific feature associated with a particular primer set, especially given that dozens of other
377 metabarcoding studies have identified similar patterns (69, 70, 100).

378 This phenomenon of stochastic dropouts adds noise to the observations and limits the
379 accuracy with which we might predict amplicon abundances (particularly rare ones). This is
380 best visualized by the noise near the origin at Figures S3 & S4. To address this, we developed a
381 comprehensive joint Bayesian model that incorporates stochasticity in observed amplicon read

382 counts through a multinomial subsampling process (See supplement 2 for full model
383 description). Thus, we explicitly account for stochasticity in the model through sampling
384 distributions and using the resulting parameters to estimate the uncertainty around our given
385 estimated larvae counts. Ultimately, such noise in the dataset does not fundamentally change
386 the interpretation of our observations or of our model but serves to limit our confidence in the
387 abundance of rare targets, a persistent problem in community ecology (102).

388

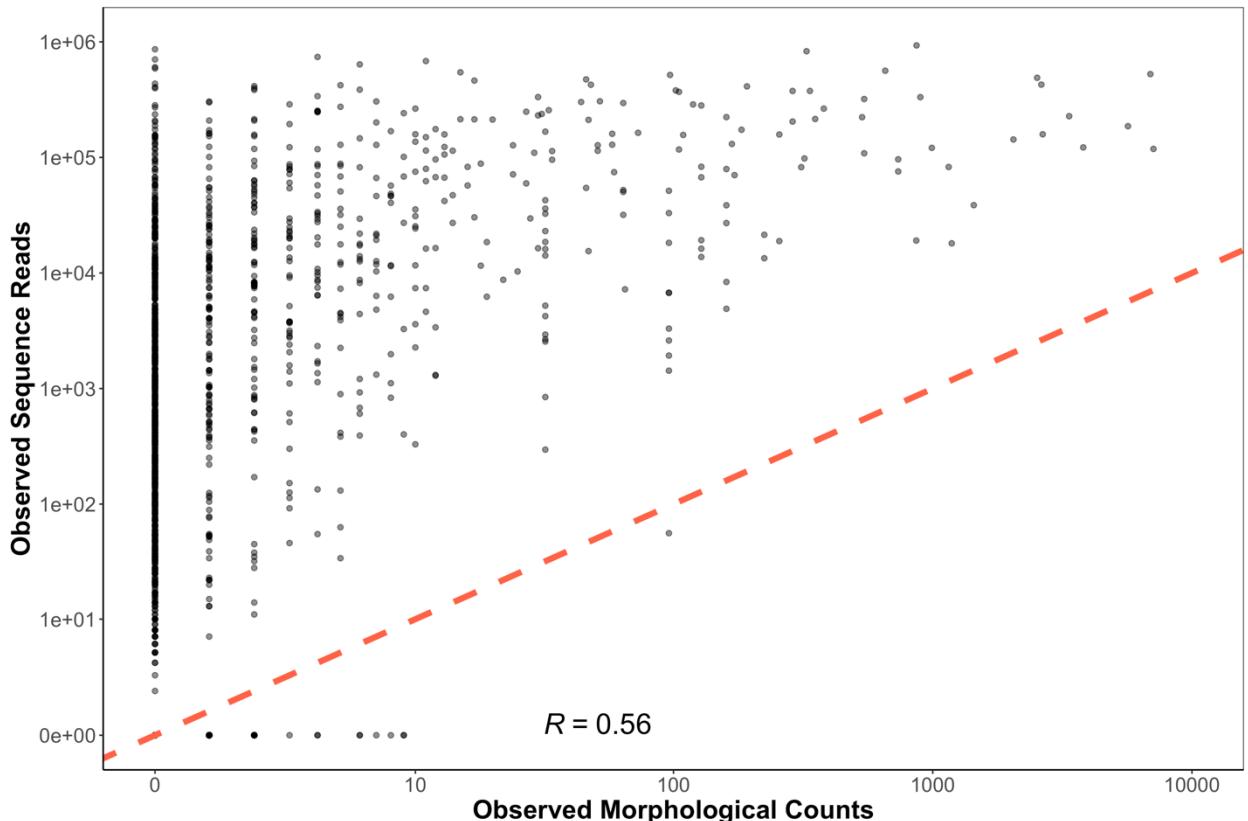
389 **Supplemental Figures**



390

391 **Figure S1. station Map**

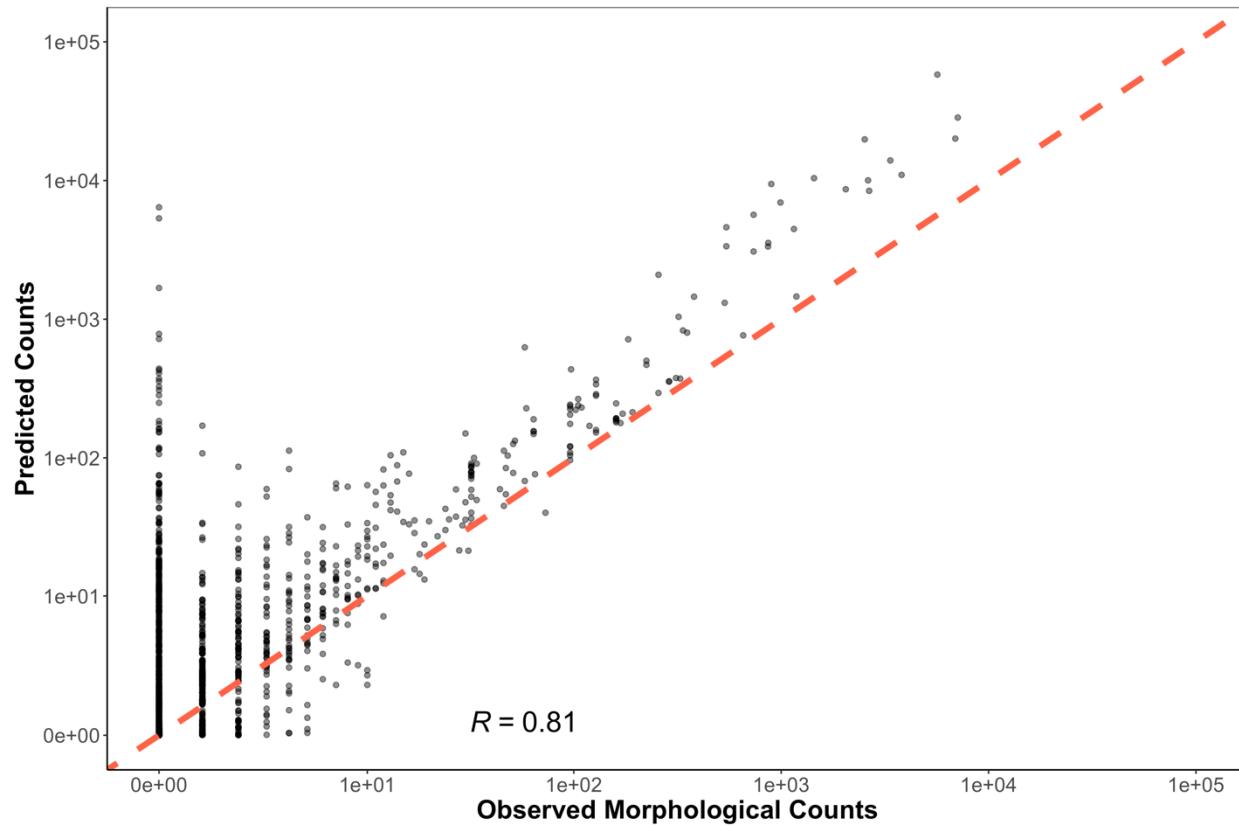
392 Ichthyoplankton samples were collected from four stations with distinct biogeographic
393 characteristics.



394

395 Figure S2. Observed Sequence Reads vs. Observed Morphological Counts

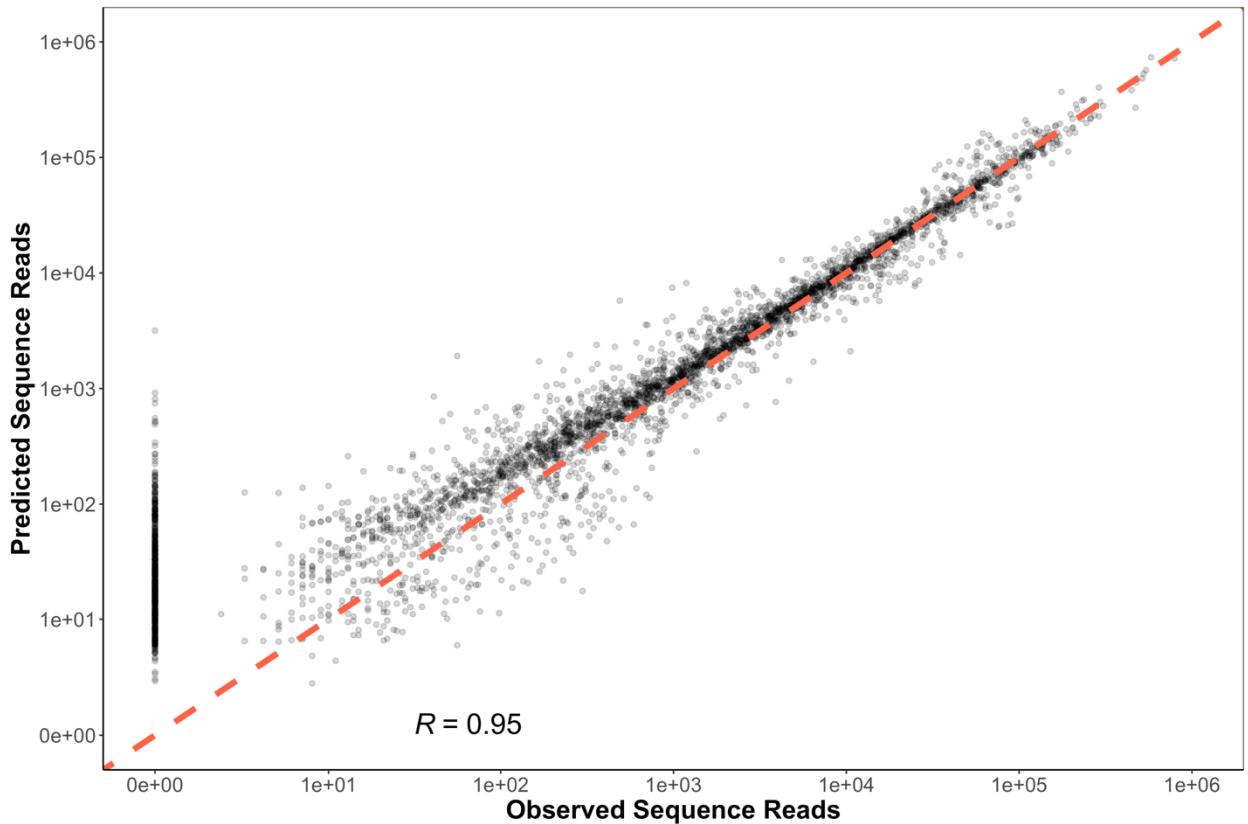
396 Observed sequencing reads and morphological counts do not follow a clear linear
 397 relationship. The one-to-one line is plotted in red and Pearson correlation coefficient is
 398 0.56. This non-linearity is unsurprising given that observed reads are a function of both
 399 DNA concentrations (here assumed proportional to morphological counts) as well as
 400 species-specific amplification efficiencies (here are unknown) (See Supplement 2). Thus
 401 without accounting for species-specific amplification efficiencies within the compositional
 402 amplicon data set, we do not expect to observe a clear positive relationship. This apparent
 403 lack of relationship depicted here motivated the creation of the mechanistic joint Bayesian
 404 model. The occurrence of stochastic dropouts (technical replicates with zero reads) can be
 405 observed along the X-axis. Variance is highest at low observed morphological counts.



406

407 Figure S3. Predicted Counts vs. Observed Morphological Counts

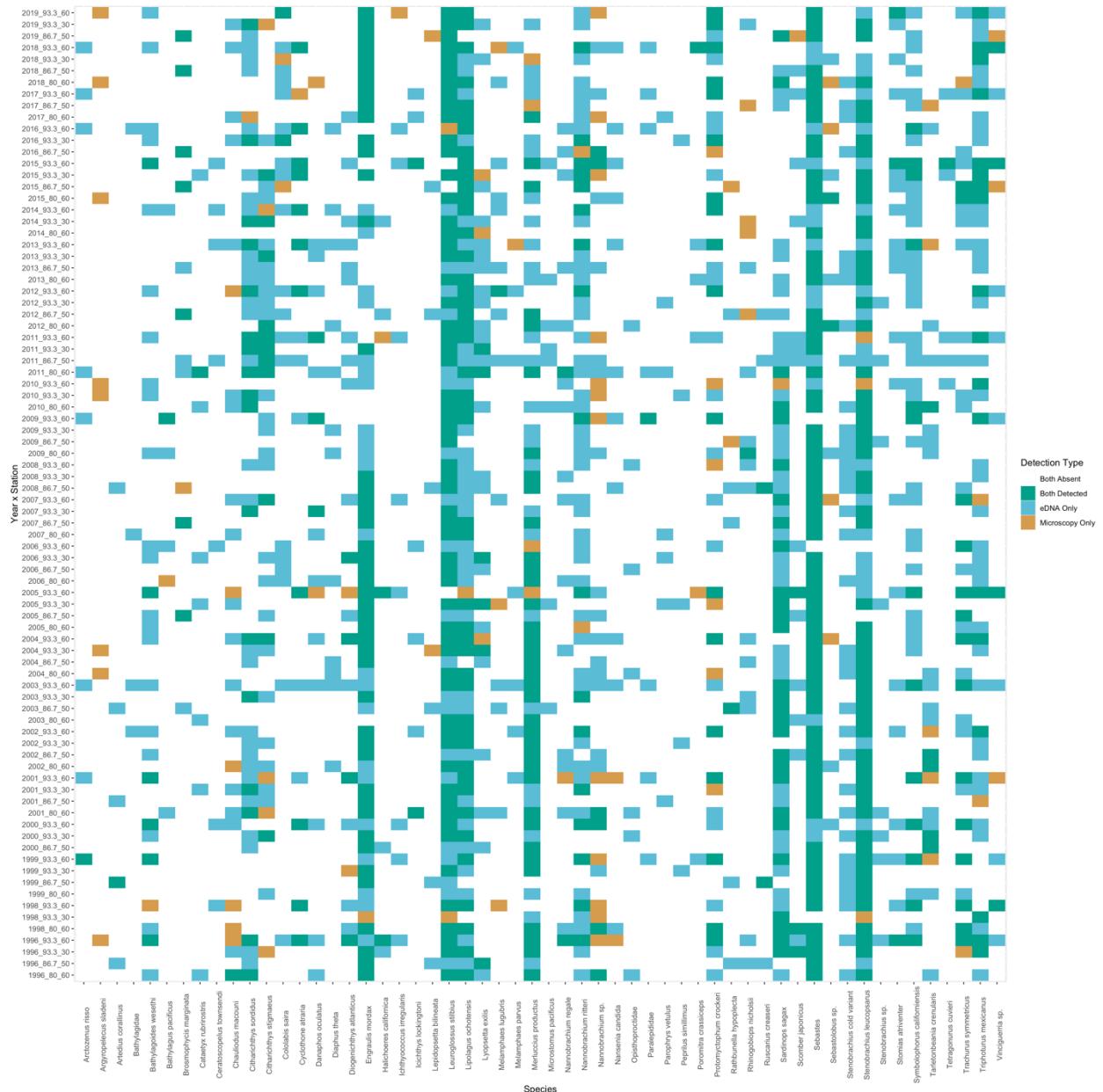
408 Predicted counts are generated from the joint Bayesian model. The one-to-one line is
 409 plotted in red and Pearson correlation coefficient is 0.81. Variance in predicted counts
 410 markedly decreases with higher observed morphological counts. We note variance in
 411 predicted counts is substantially less than that between observed reads and morphological
 412 counts Figure S3.



413

414 Figure S4. Predicted Sequence Reads vs. Observed Sequence Reads

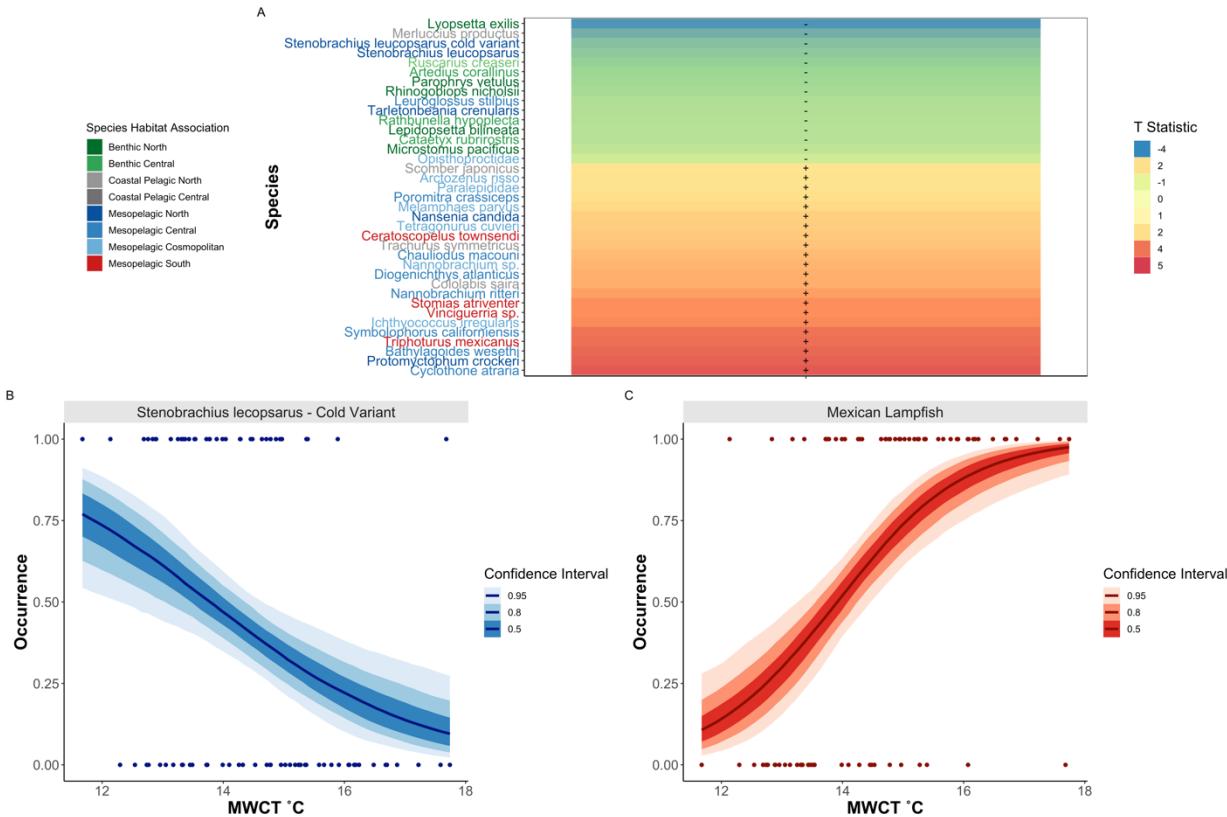
415 Predicted sequence reads are generated from the joint Bayesian model. The one-to-one
 416 line is plotted in red and Pearson correlation coefficient is 0.95. In general, predicted
 417 sequence reads track observed sequence reads, and show substantially less variance than
 418 observed sequence reads in Figure S3. However, unexpected zeros across multiple
 419 technical PCR replicates (stochastic dropouts), deviate notably from expected, low
 420 variance results (points along Y-axis). We note that all observed stochastic dropouts occur
 421 in less than 2.9% of sample read proportions and less than 9 morphological counts for a
 422 given sample. Such dropouts are likely a function of subsampling rare DNA molecules
 423 associated with molecular biology processing (See Supplement 2).



424

425 **Figure S5. Co-detection of Taxa By Metabarcoding and Microscopy**

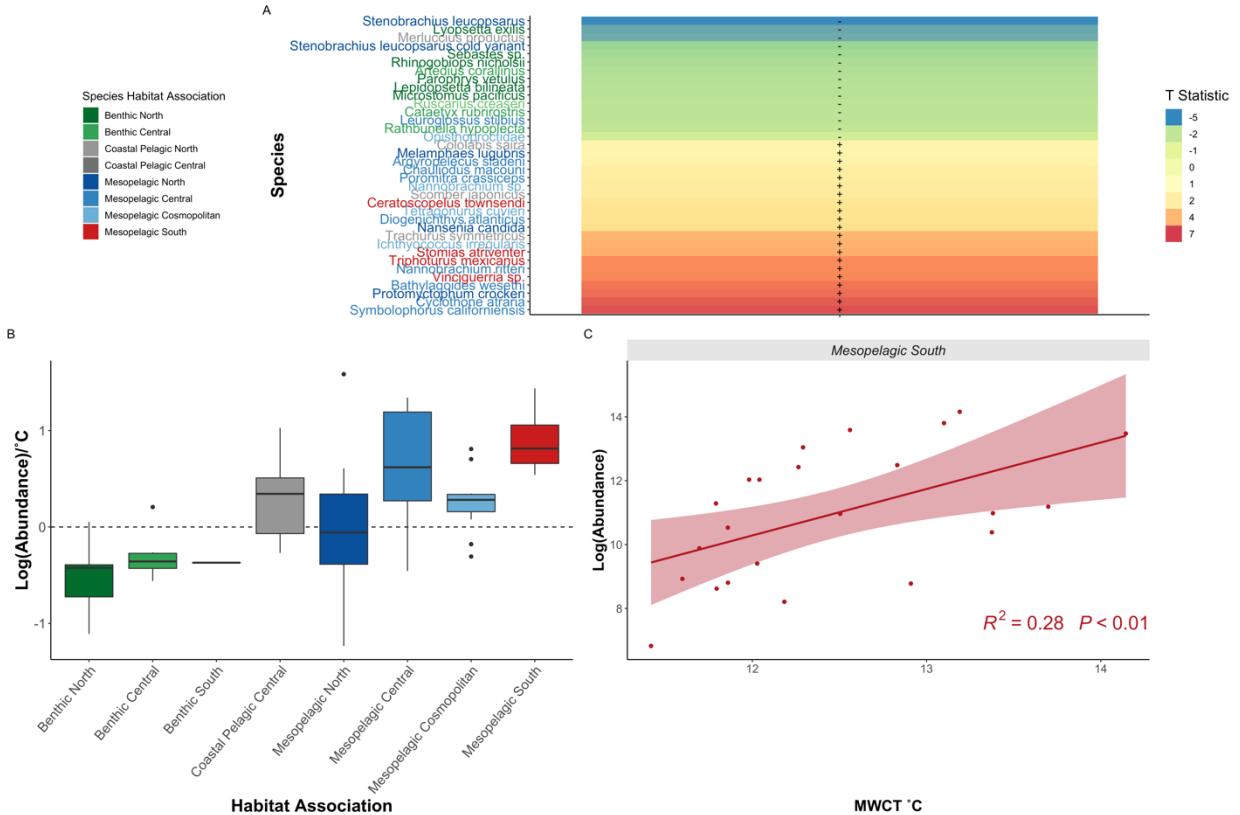
426 Of the 56 taxa used for modeling efforts (Supplemental Methods), both metabarcoding
 427 and microscopy detected 46 taxa, with nine detected only by metabarcoding and one
 428 detected only by microscopy.



429

430 **Figure S6. Temperature Associations in Fish Species**

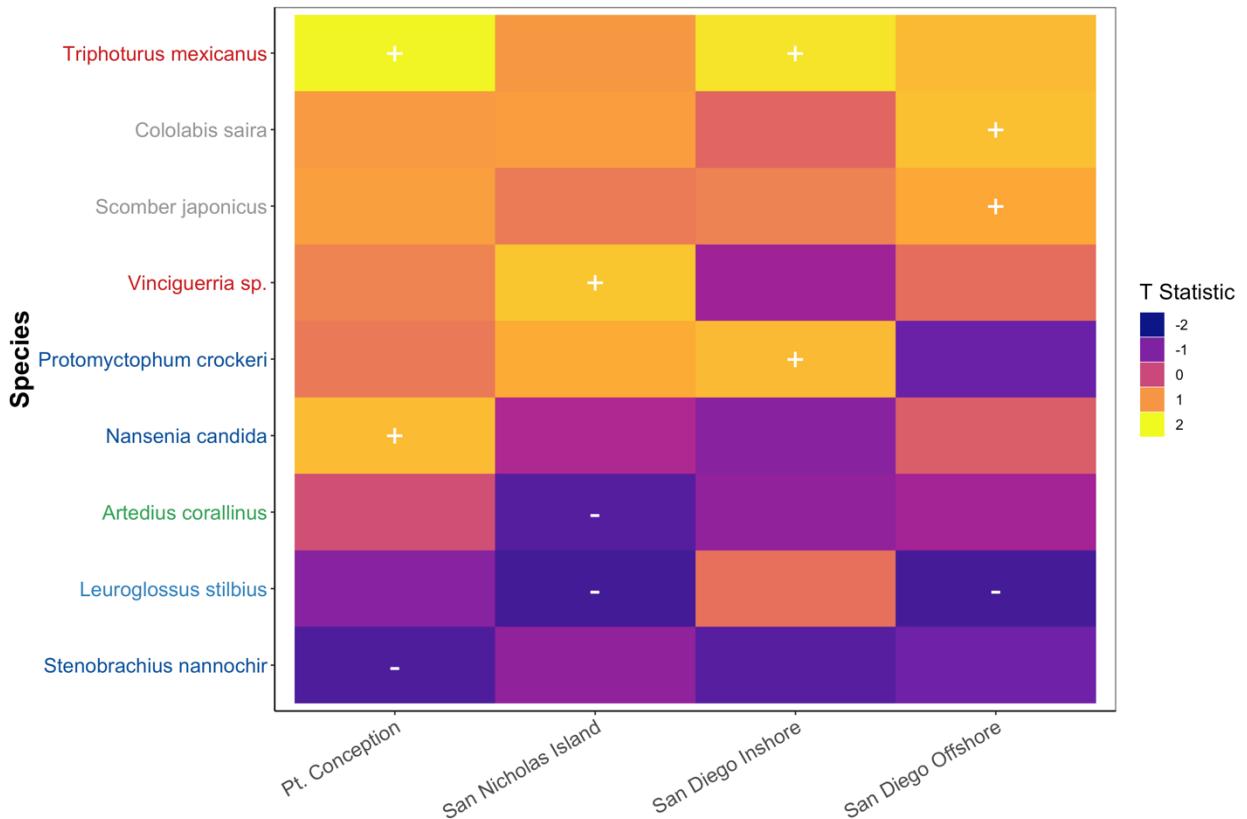
431 Changes in species occurrence patterns in response to MWCT, with Southern
 432 Mesopelagic species increasing in prevalence with elevated temperature (A). T statistic
 433 (slope coefficient/ standard error) from generalized binomial mixed model was calculated
 434 for each species across all stations. Only species with significantly different slopes (95%
 435 CI greater or less than zero) are plotted. Importantly, metabarcoding identified cold
 436 associated variants of the Northern Lanternfish (*Stenobrachius leucopsarus*) which
 437 cannot be morphologically identified (B) as well as warm-associated species like the
 438 Mexican Lampfish (*Triphoturus mexicanus*) (C).



439

440 **Figure S7. Southern Oceanic Species Drive Fish Community Shifts**

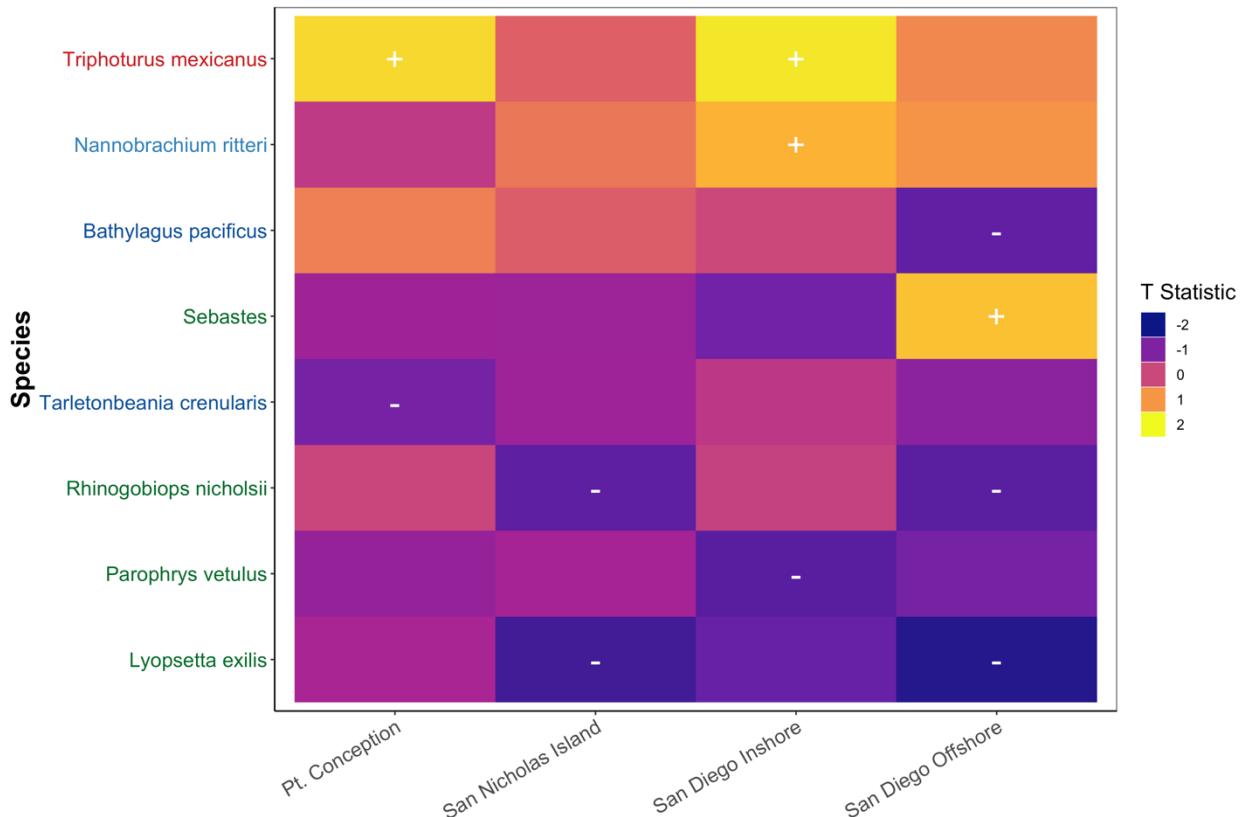
441 We capture changes in species biomass in response to MWCT, with Southern
 442 Mesopelagic species increasing in abundance with elevated temperature (A). T statistic
 443 (slope coefficient/standard error) from generalized linear models were calculated for each
 444 species across all stations. Only species with significantly different slopes (95% CI
 445 greater or less than zero) are plotted. Southern mesopelagic fishes were associated with
 446 increased temperature as indicated by the boxplots of all species-specific slopes from
 447 generalized linear models (B) and by the aggregated abundance relationship (C). In
 448 contrast, benthic species, as well as Northern Hake and Pacific Sardine abundances, were
 449 correlated with cooler temperatures.



450

451 **Figure S8. Significant Species Occurrence and SST Correlations at each Station**

452 Occurrence of the Mexican lampfish (*Triphoturus mexicanus*) was positively correlated
 453 with increased SST at the three northernmost stations. Generalized binomial mixed model
 454 of occurrence versus SST was calculated for each species at each station. Only species
 455 with significantly different slopes (95% CI greater or less than zero) are plotted. Colors
 456 correspond to T statistic (slope coefficient/ standard error). Colors of species names
 457 correspond to habitat associations described in Figure 2.



458

459

Figure S9. Significant Species Occurrence and MWCT Correlations at each Station

460

Occurrence of the Mexican lampfish (*Triphoturus mexicanus*) was positively correlated

461

with increased MWCT at the three northernmost stations. Generalized binomial mixed

462

model of occurrence versus MWCT was calculated for each species at each station. Only

463

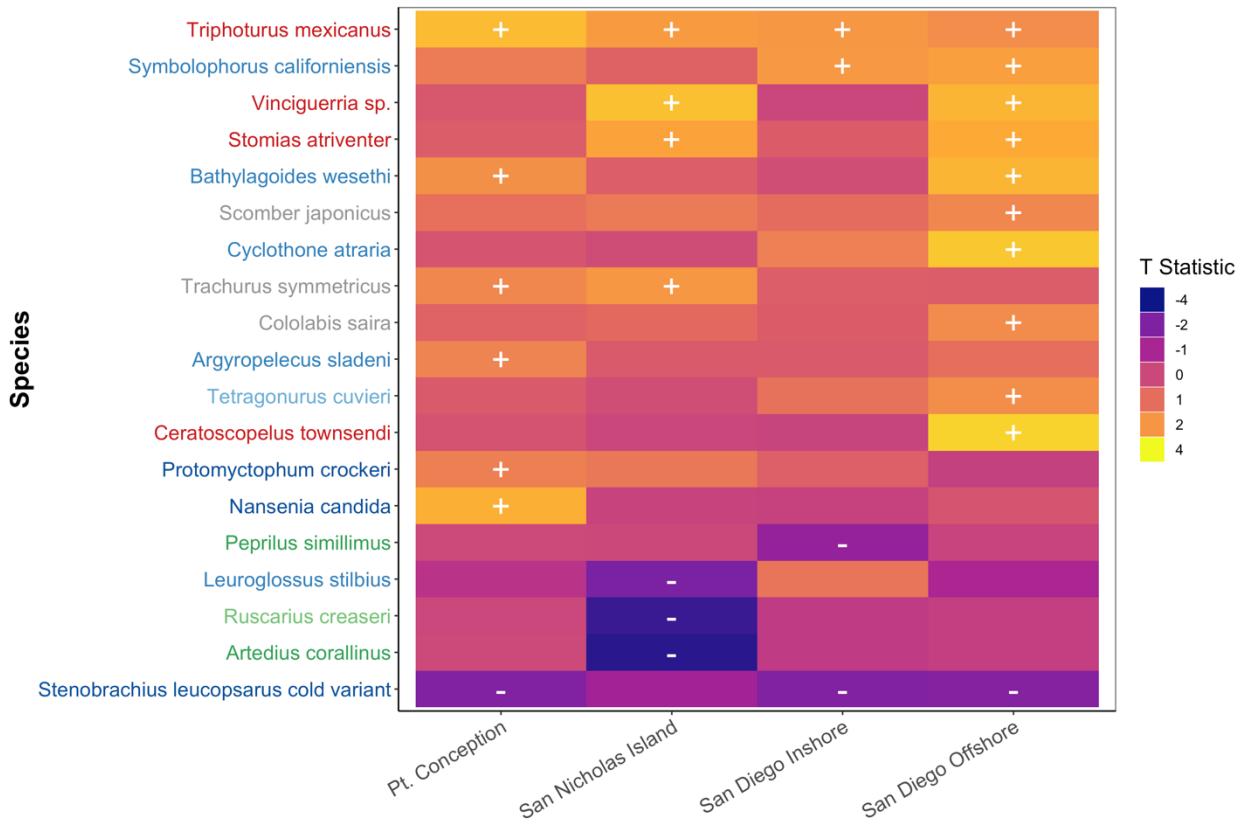
species with significantly different slopes (95% CI greater or less than zero) are plotted.

464

Colors correspond to T_statistic (slope coefficient/ standard error). Colors of species

465

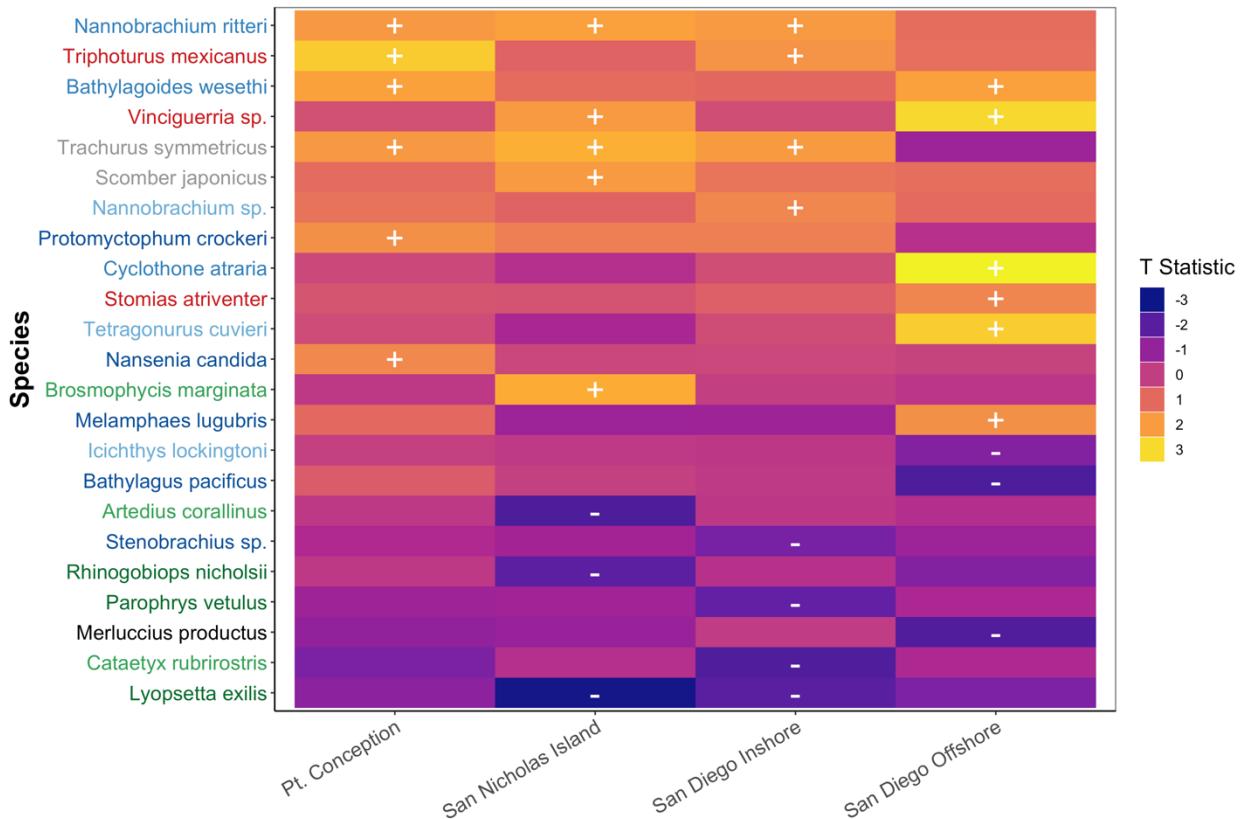
names correspond to habitat associations described in Figure 2.



466

467 **Figure S10. Significant Species Abundance and SST Correlations at each station**

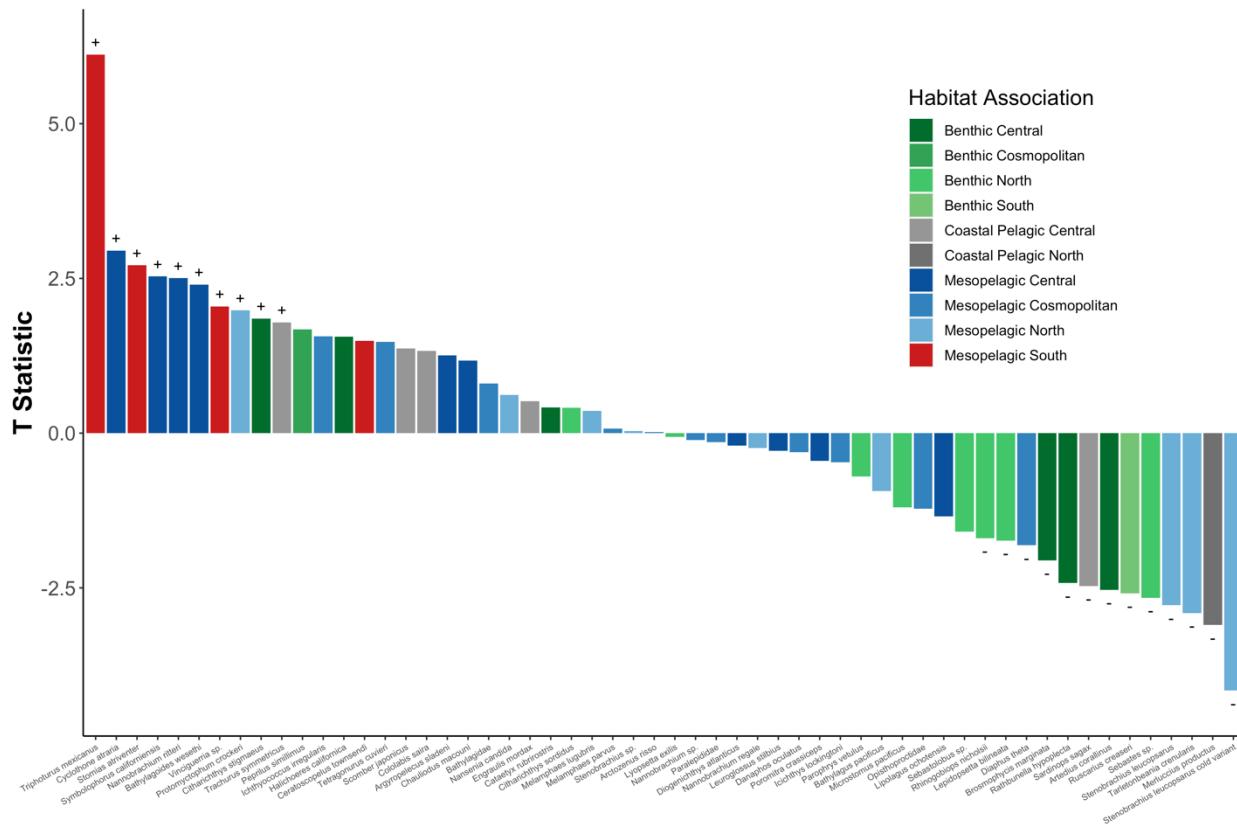
468 The abundance of the Mexican lampfish (*Triphoturus mexicanus*) was positively
 469 associated with increased SST at all stations. Likewise, the abundance of suite of
 470 mesopelagic species including *Vinciguerra* sp., *Symbolophorus californiensis*, and
 471 *Stomias atriventer* among others increased with warmer SST. Generalized linear mixed
 472 model of log (abundance) versus SST was calculated for each species at each station. Only
 473 species with significantly different slopes (95% CI greater or less than zero) are plotted.
 474 Colors correspond to T statistic (slope coefficient/ standard error). Colors of species
 475 names correspond to habitat associations described in Figure 2.



476

477 **Figure S11. Significant Species Abundance and MWCT Correlations at each station**

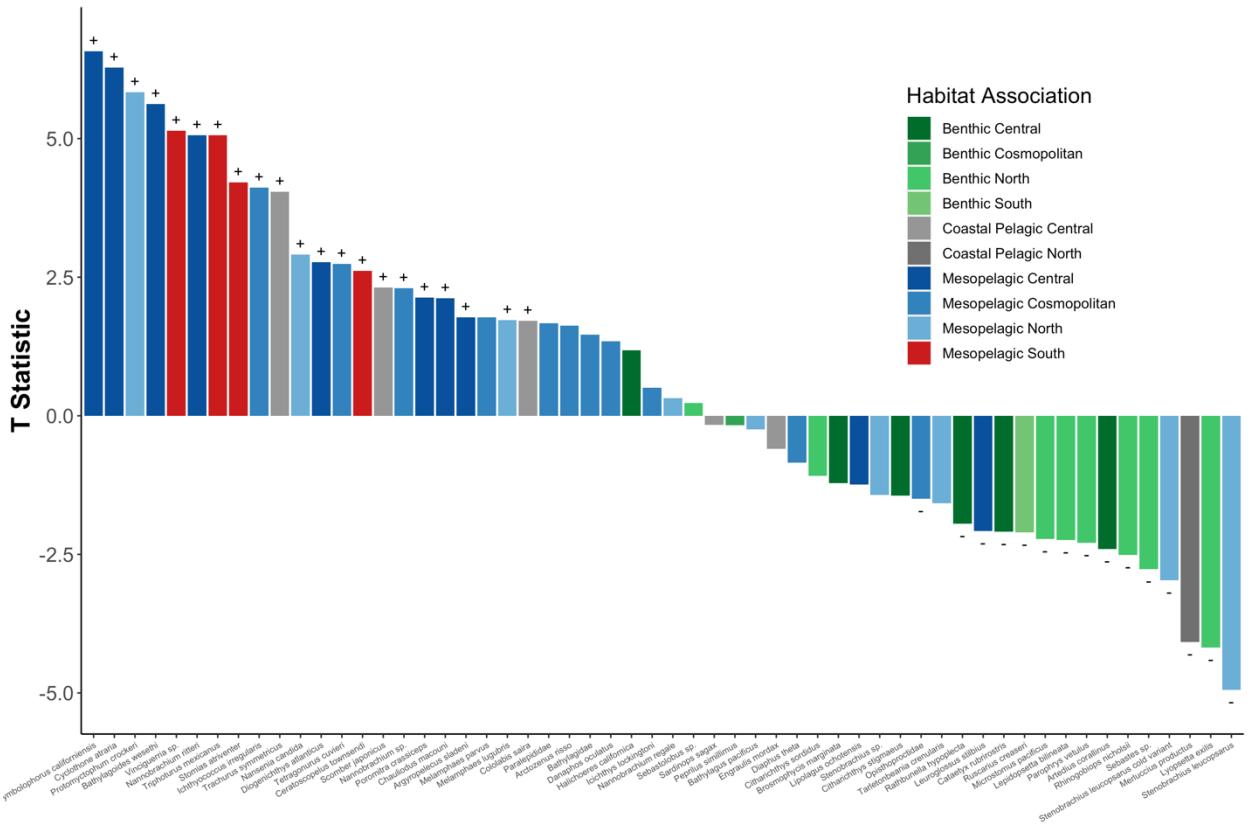
478 The abundance of the Mexican lampfish (*Triphoturus mexicanus*) was positively
 479 associated with increased MWCT at all stations. Likewise, the abundance of suite of
 480 mesopelagic species including *Vinciguerra* sp., *Symbolophorus californiensis*, and
 481 *Stomias atriventer* among others increased with warmer SST. Generalized linear mixed
 482 model of log (abundance) versus MWCT was calculated for each species at each station.
 483 Only species with significantly different slopes (95% CI greater or less than zero) are
 484 plotted. Colors correspond to T statistic (slope coefficient/ standard error). Colors of
 485 species names correspond to habitat associations described in Figure 2.



486

487 **Figure S12. Bar Plot of Significant Species Abundance and SST Correlations Across**
 488 **All stations**

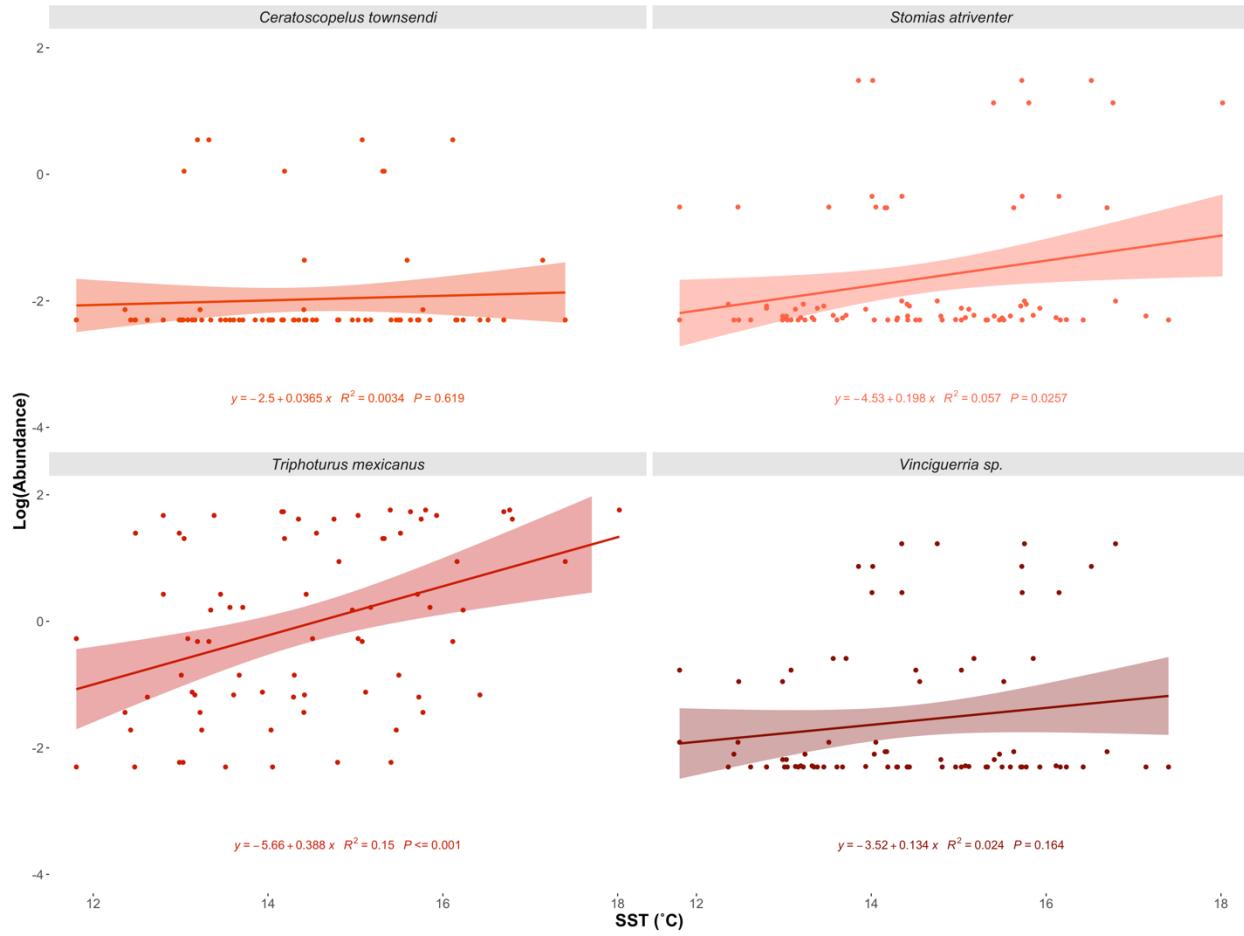
489 The abundance of southern mesopelagic species increased with warmer temperatures
 490 while fisheries targets like Pacific Sardine (*Sardinops sagax*) and North Pacific Hake
 491 (*Merluccius productus*) decreased. Generalized linear mixed model of log (abundance)
 492 versus SST was calculated for each species across all stations. Only species with
 493 significantly different slopes (95% CI greater or less than zero) are plotted. Colors
 494 correspond to T statistic (slope coefficient/ standard error).



495

496 **Figure S13. Bar Plot of Significant Species Abundance and MWCT Correlations**497 **Across All stations**

498 The abundance of southern mesopelagic species increased with warmer temperatures
 499 while fisheries targets like Pacific Sardine (*Sardinops sagax*) and North Pacific Hake
 500 (*Merluccius productus*) decreased. Generalized linear mixed model of log (abundance)
 501 versus SST was calculated for each species across all stations. Only species with
 502 significantly different slopes (95% CI greater or less than zero) are plotted. Colors
 503 correspond to T statistic (slope coefficient/ standard error).



504

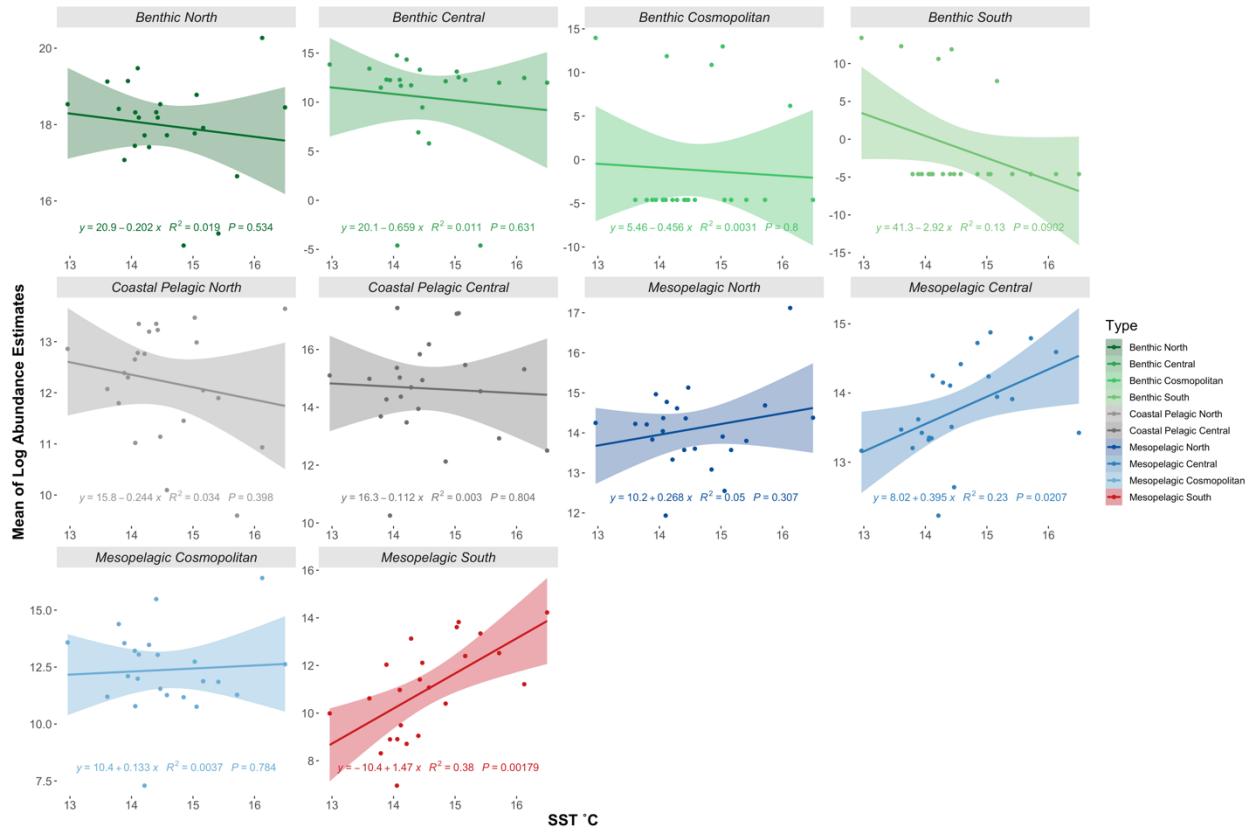
505 **Figure S14. Increased Abundance of Southern Mesopelagic Species with Higher**506 **SST**

507 Species specific regressions of log (Abundance) vs. SST (°C). Three of the four southern

508 mesopelagic species had significant positive associations between abundance and SST.

509 Although not significant across all stations, the abundance of *Ceratoscopelus townsendi*

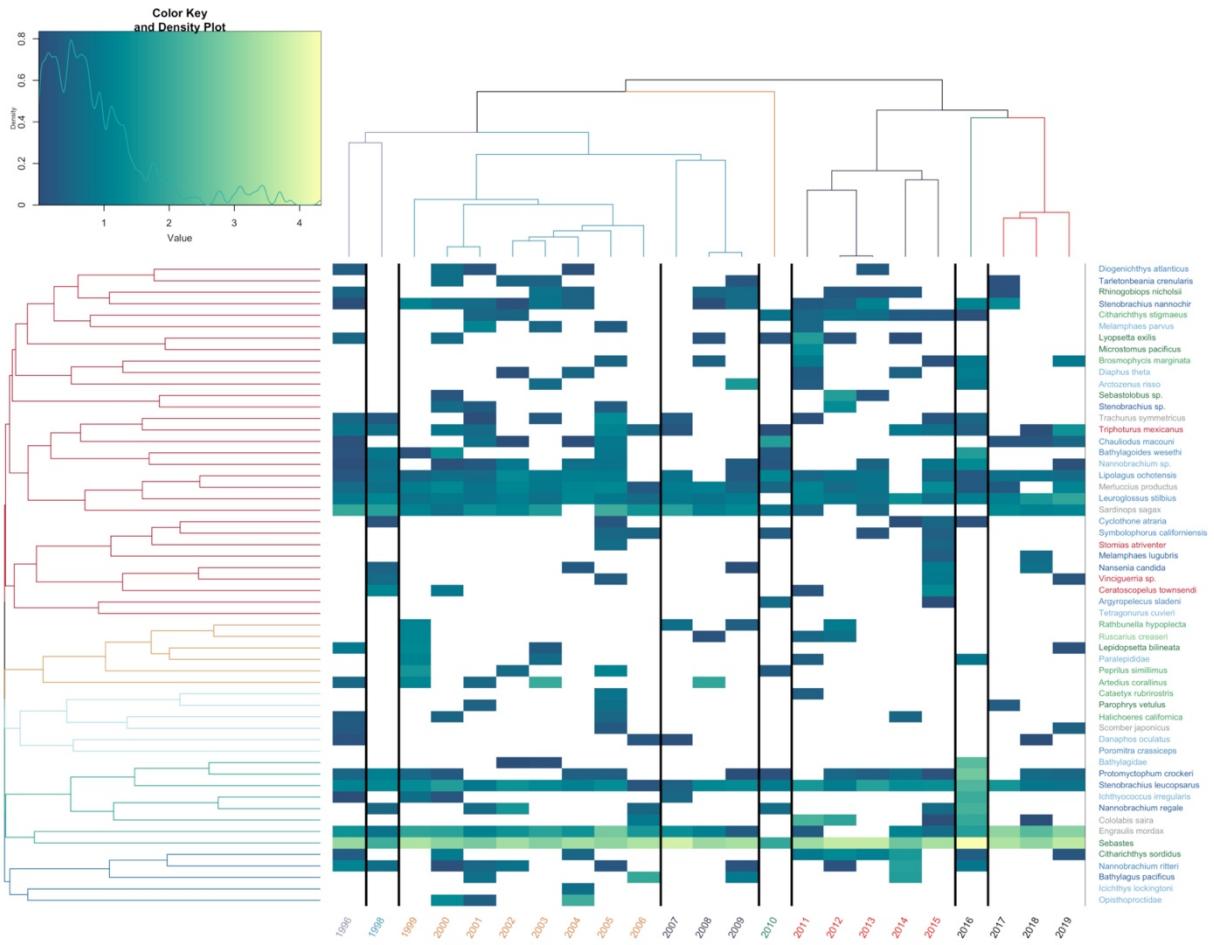
510 significantly increased with higher SST at the San Diego Offshore station (Figure S7).



511

512 **Figure S15. Increased Abundance of Southern and Central Mesopelagic Species with
513 Higher SST**

514 Abundances of each habitat association were summed at each station. Habitat association
515 regressions were fit using sum total log (Abundance) vs. SST (°C).



516

517

Figure S16. Heat Map of Abundances Over Time

518

Northern Anchovy (*Engraulis mordax*) and rockfishes *Sebastodes* sp. dominated predicted

519

counts. Estimated abundance of each year, averaged across stations, plotted over time.

520

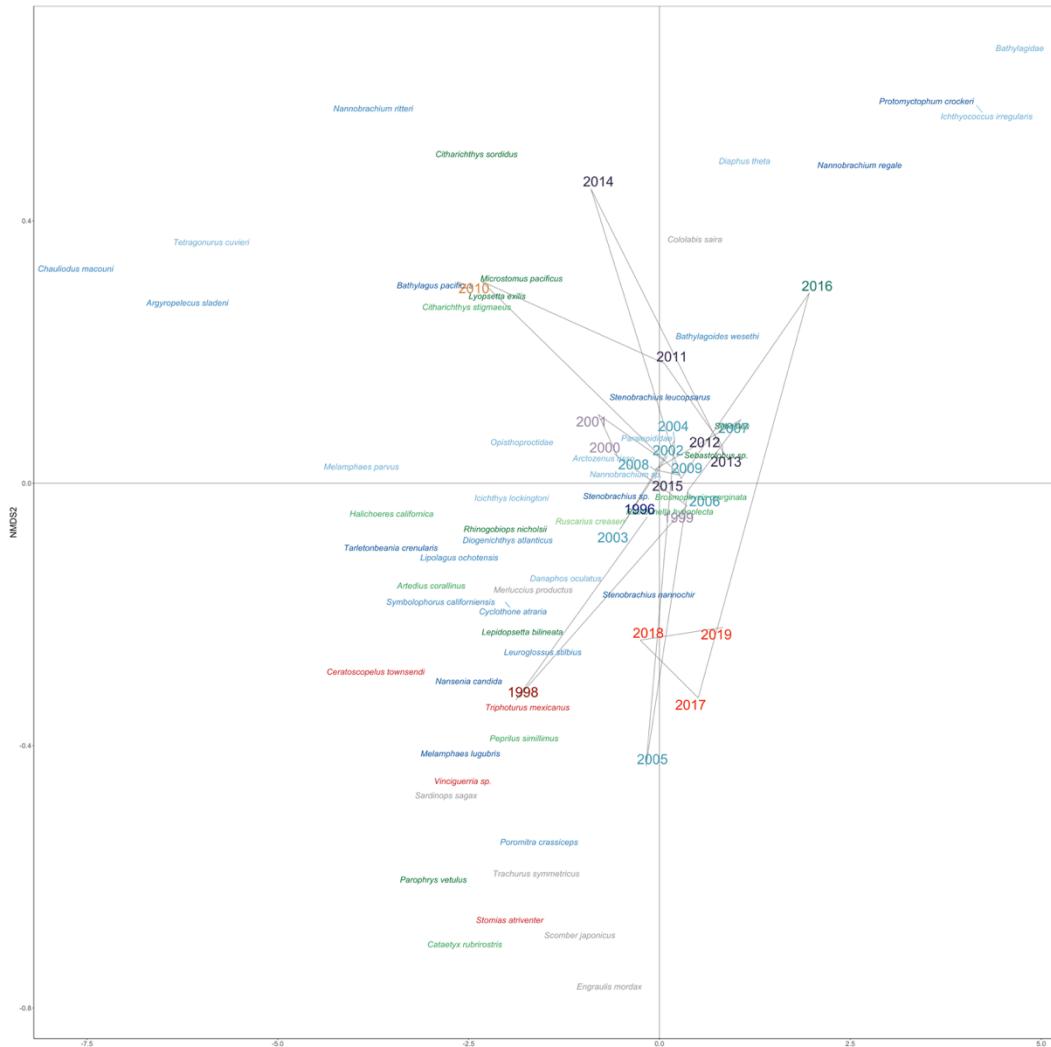
Years are color coded by chronological clustering. Species are grouped by hierarchical

521

clustering. Lighter colors indicate higher abundance, white is a lack of detection. Species

522

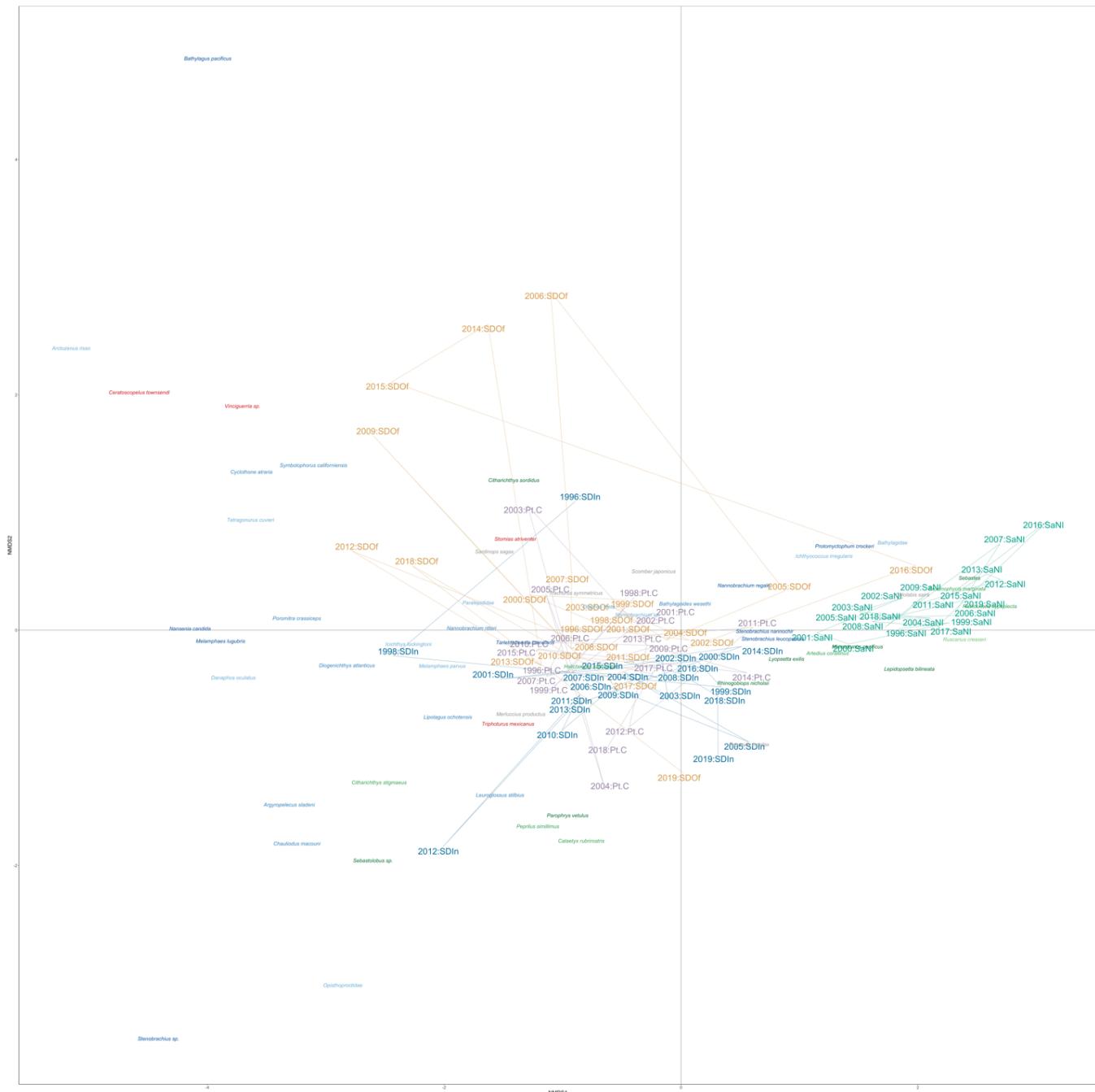
are color coded by habitat association matching Figure 1.



523

524 **Figure S17. NMDS Ordination of Species and Years**

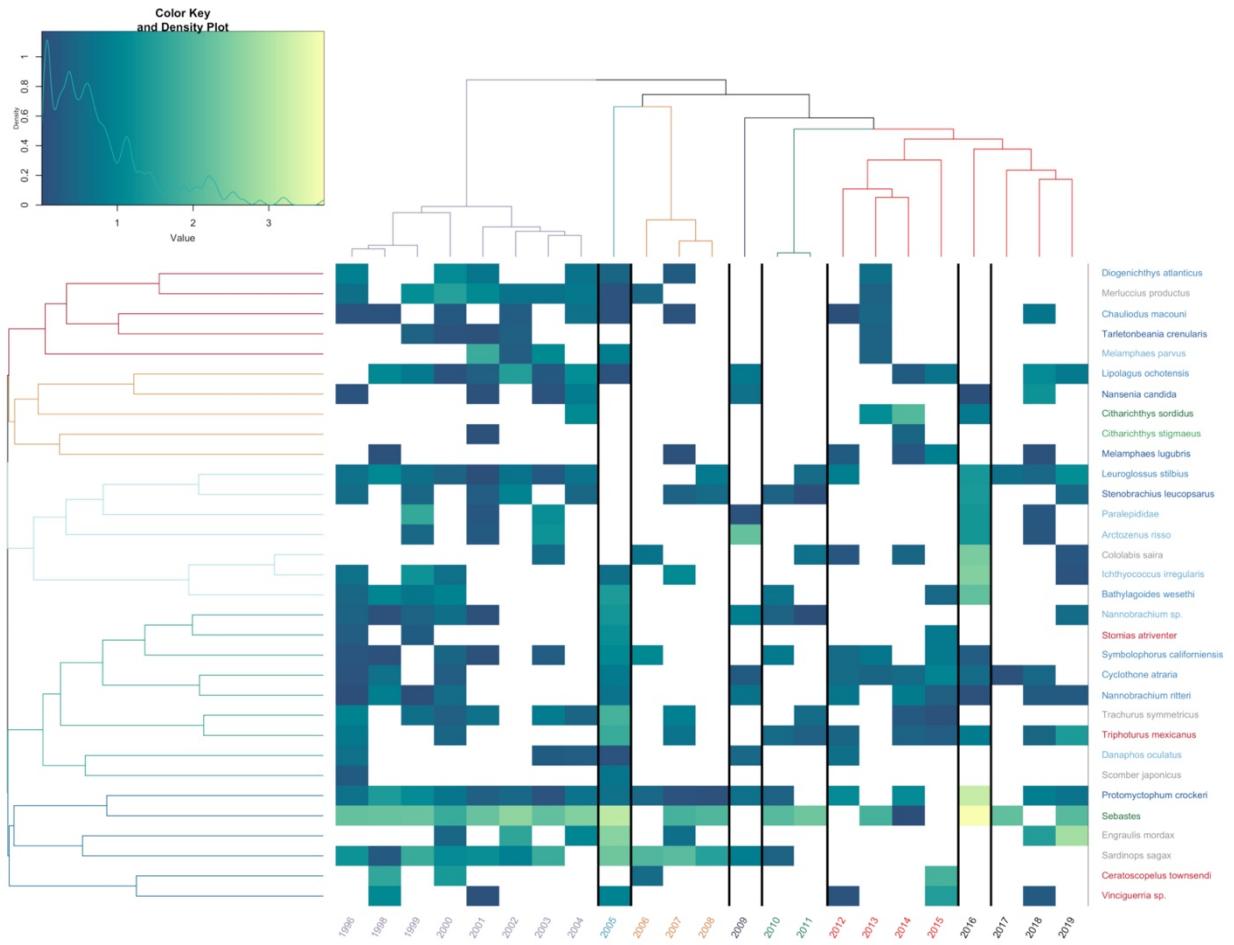
525 Fish assemblages changed across time with southern mesopelagics clustering with the
 526 1998 and 2005 El Niños as well as 2017-2019 after the peak of the marine heat wave.
 527 NMDS Ordination of Bray-Curtis dissimilarities calculated from summed abundance of
 528 each year averaged across stations. Marine heatwave patterns are obscured by differential
 529 onset and receding of the warming event across stations. Years are color coded by
 530 chronological clustering ($k=8$). Species are color coded by habitat association matching
 531 Figure 2.



532

533 **Figure S18. NMDS Ordination of Species and Samples**

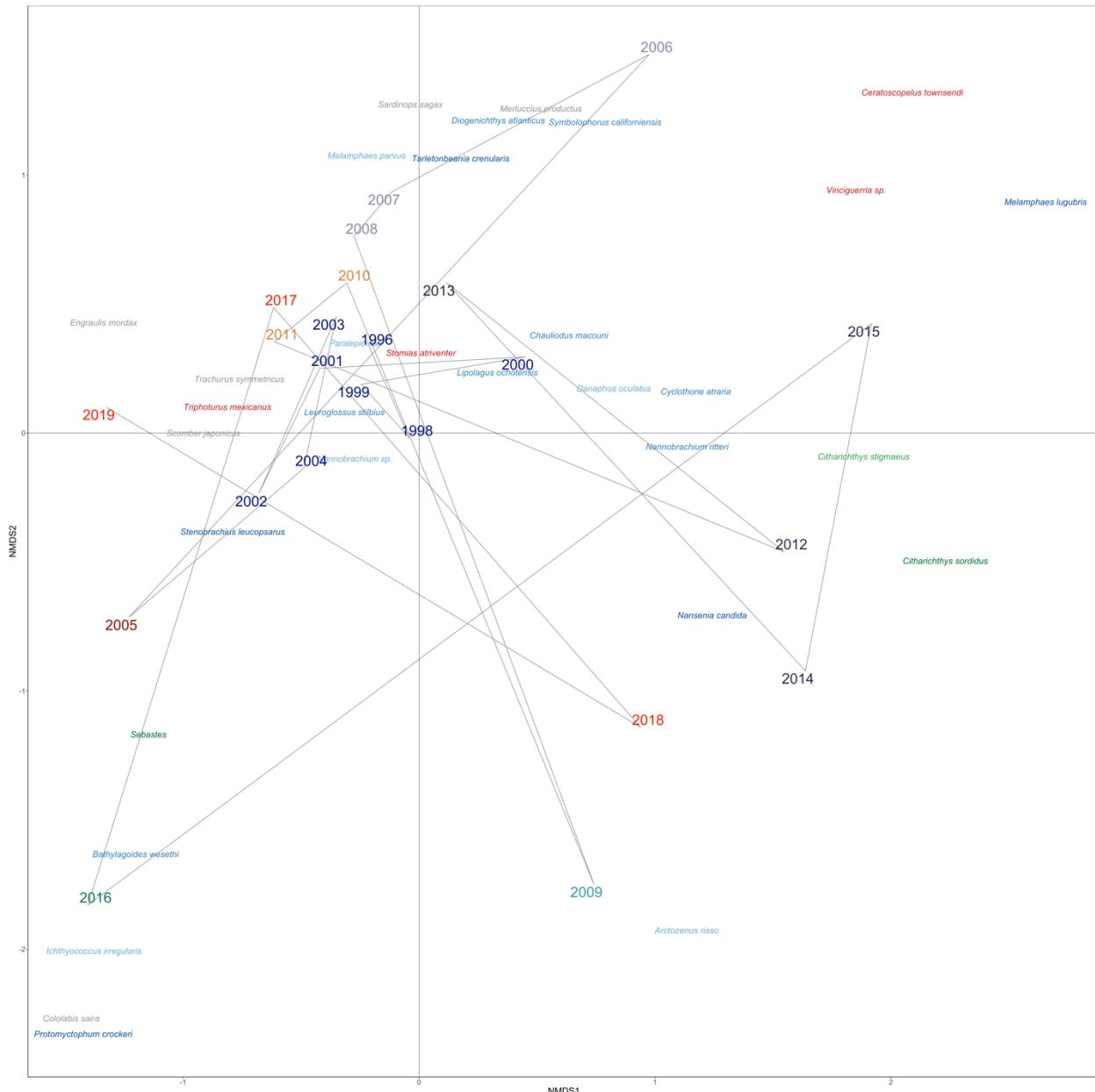
534 Fish assemblages were strongly structured by stations, particularly the San Nicholas Island
 535 station which had the highest abundance of *Sebastes* sp. NMDS Ordination of Bray-Curtis
 536 dissimilarities calculated from the abundance of each year and station. Samples are color
 537 coded by station. Species are color coded by habitat association matching Figure 2.



538

539 **Figure S19. Heat Map of San Diego Offshore Abundances Over Time**

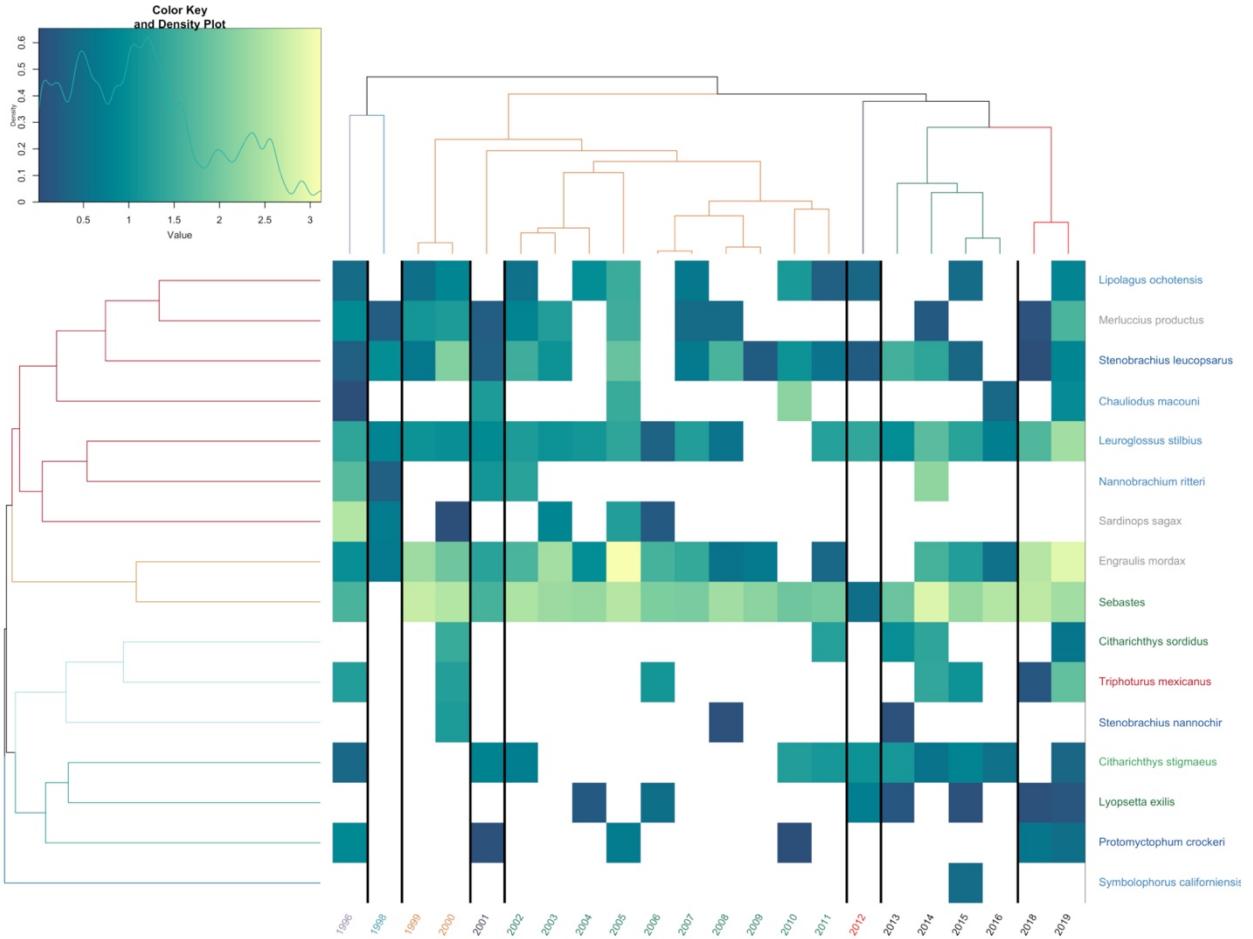
540 Estimated abundance of each year plotted over time. Years are color coded by
 541 chronological clustering. Species are grouped by hierarchical clustering. Lighter colors
 542 indicate higher abundance, white is a lack of detection. Species are color coded by habitat
 543 association matching Figure 1.



544

545 **Figure S20. NMDS Ordination of San Diego Offshore Species and Years**

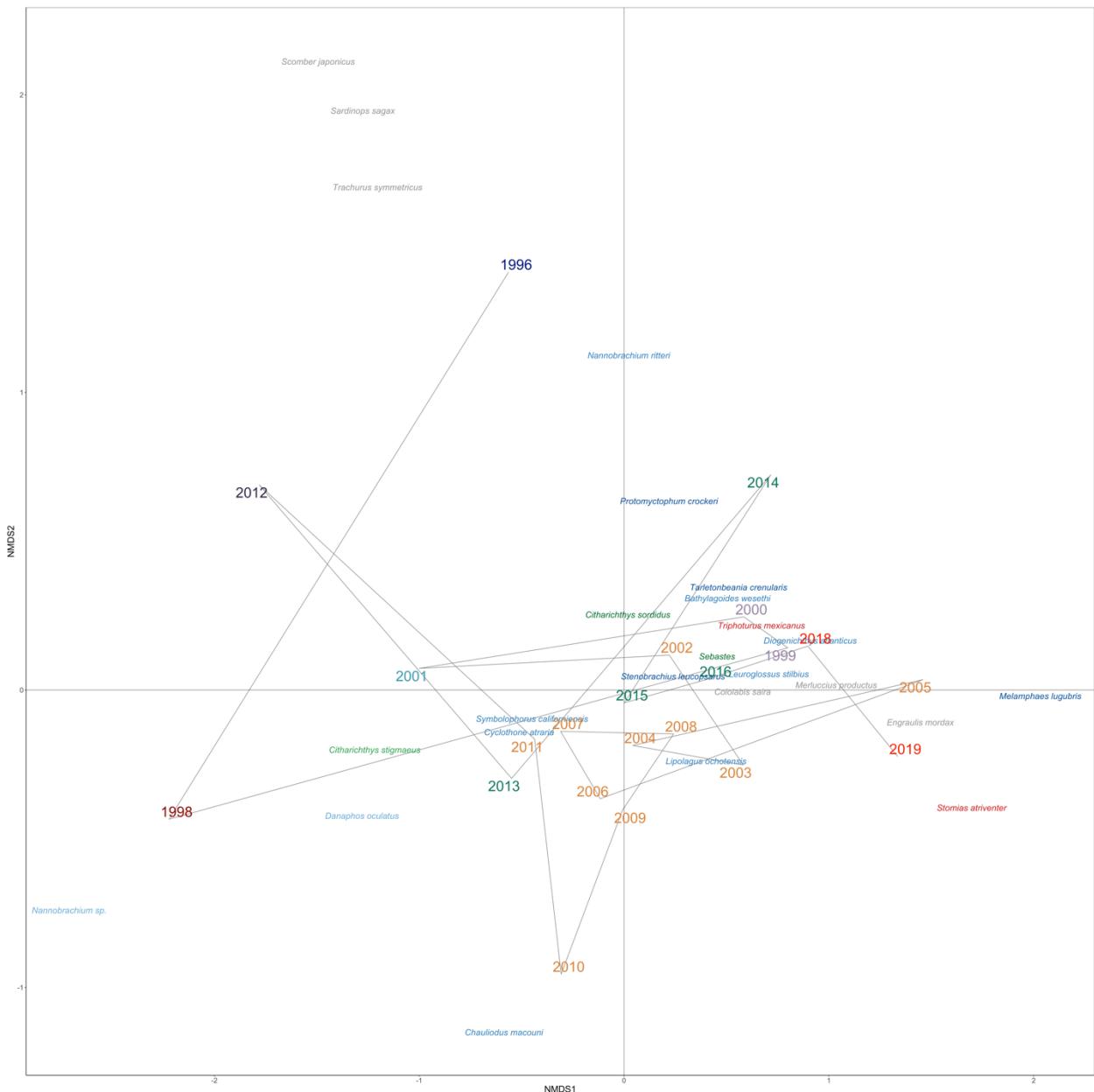
546 NMDS Ordination of Bray-Curtis dissimilarities calculated from abundance of each year.
 547 Years are color coded by chronological clustering (k =8). Species are color coded by
 548 habitat association matching Figure 2.



549

550 **Figure S21. Heat Map of San Diego Inshore Abundances Over Time**

551 Estimated abundance of each year plotted over time. Years are color coded by
 552 chronological clustering. Species are grouped by hierarchical clustering. Lighter colors
 553 indicate higher abundance, white is a lack of detection. Species are color coded by habitat
 554 association matching Figure 1.

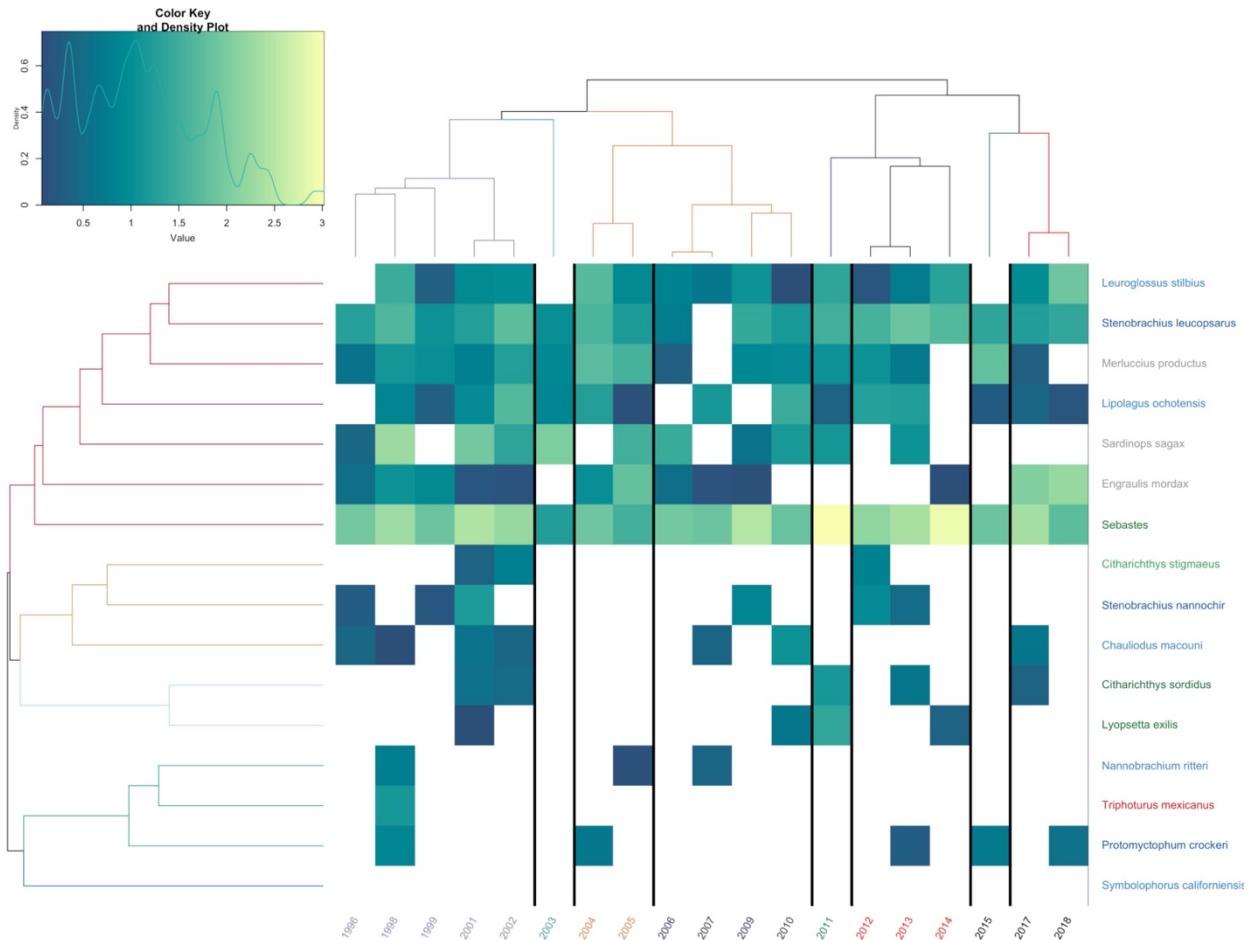


555

556 **Figure S22. NMDS Ordination of San Diego Inshore Species and Years**

557 NMDS Ordination of Bray-Curtis dissimilarities calculated from abundance of each year.

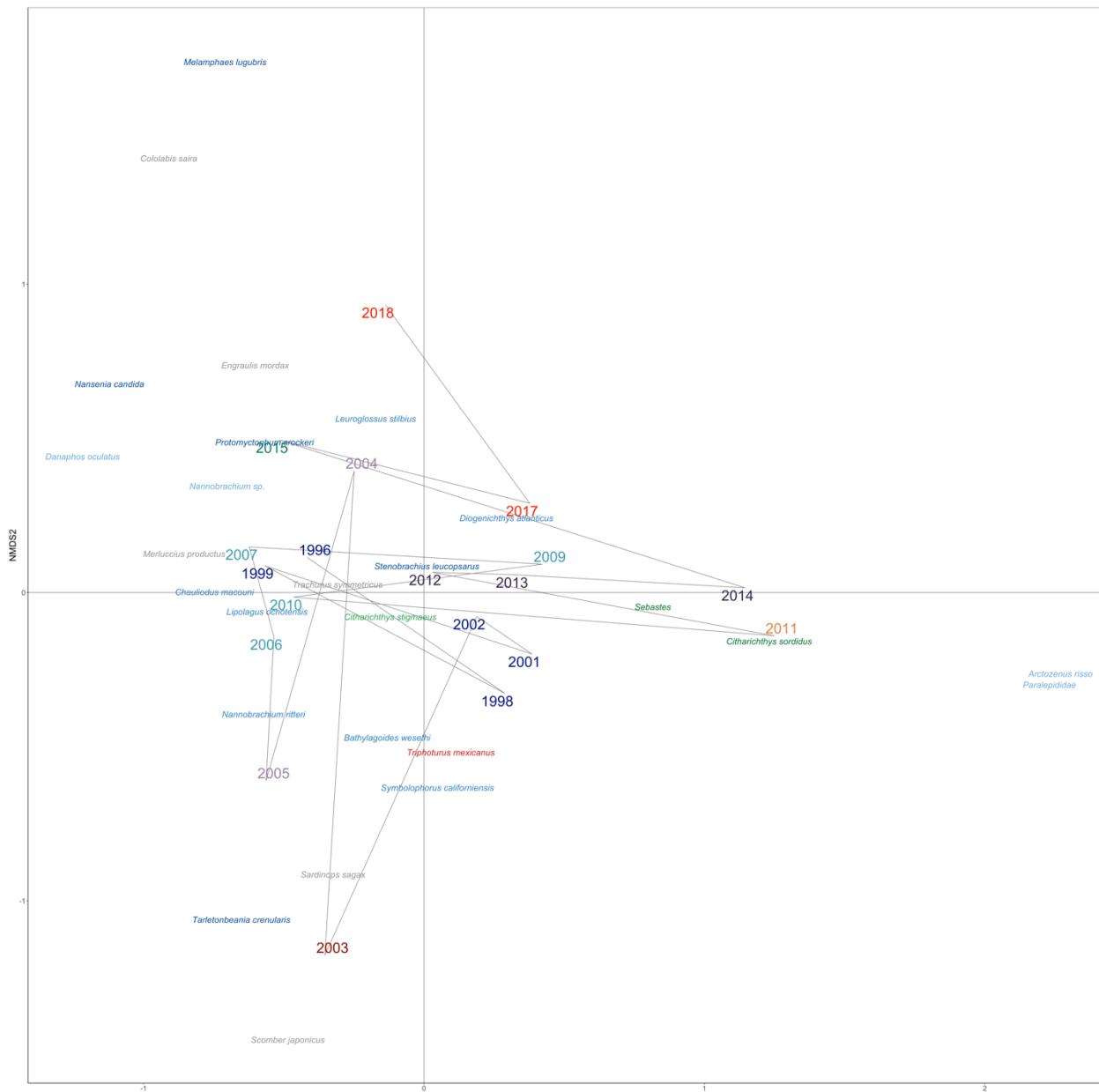
558 Years are color coded by chronological clustering ($k=8$). Species are color coded by
559 habitat association matching Figure 2.



560

561 **Figure S23. Heat Map of Pt. Conception Abundances Over Time**

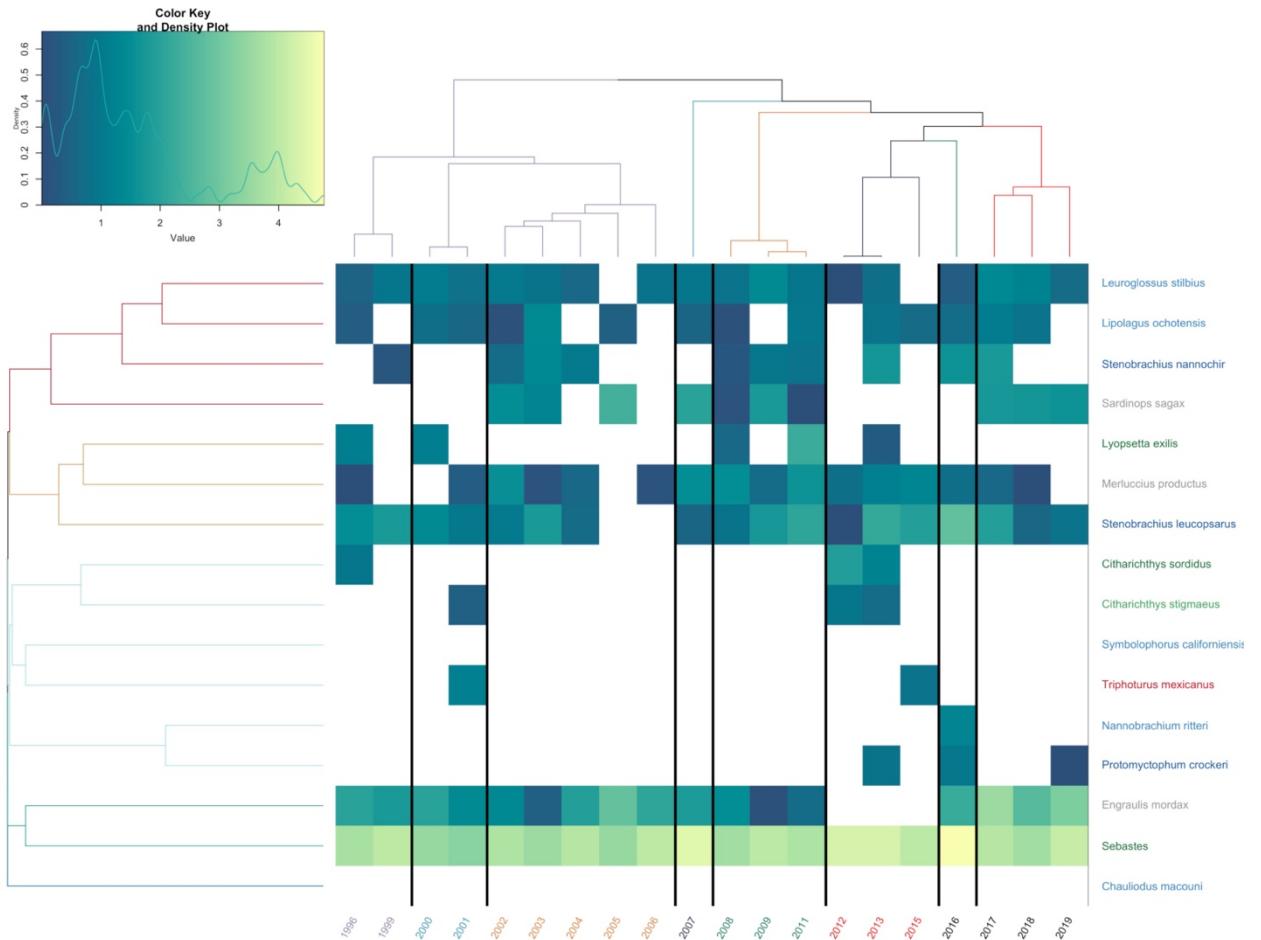
562 Estimated abundance of each year plotted over time. Years are color coded by
 563 chronological clustering. Species are grouped by hierarchical clustering. Lighter colors
 564 indicate higher abundance, white is a lack of detection. Species are color coded by habitat
 565 association matching Figure 1.



566

567 **Figure S24. NMDS Ordination of Pt. Conception Species and Years**

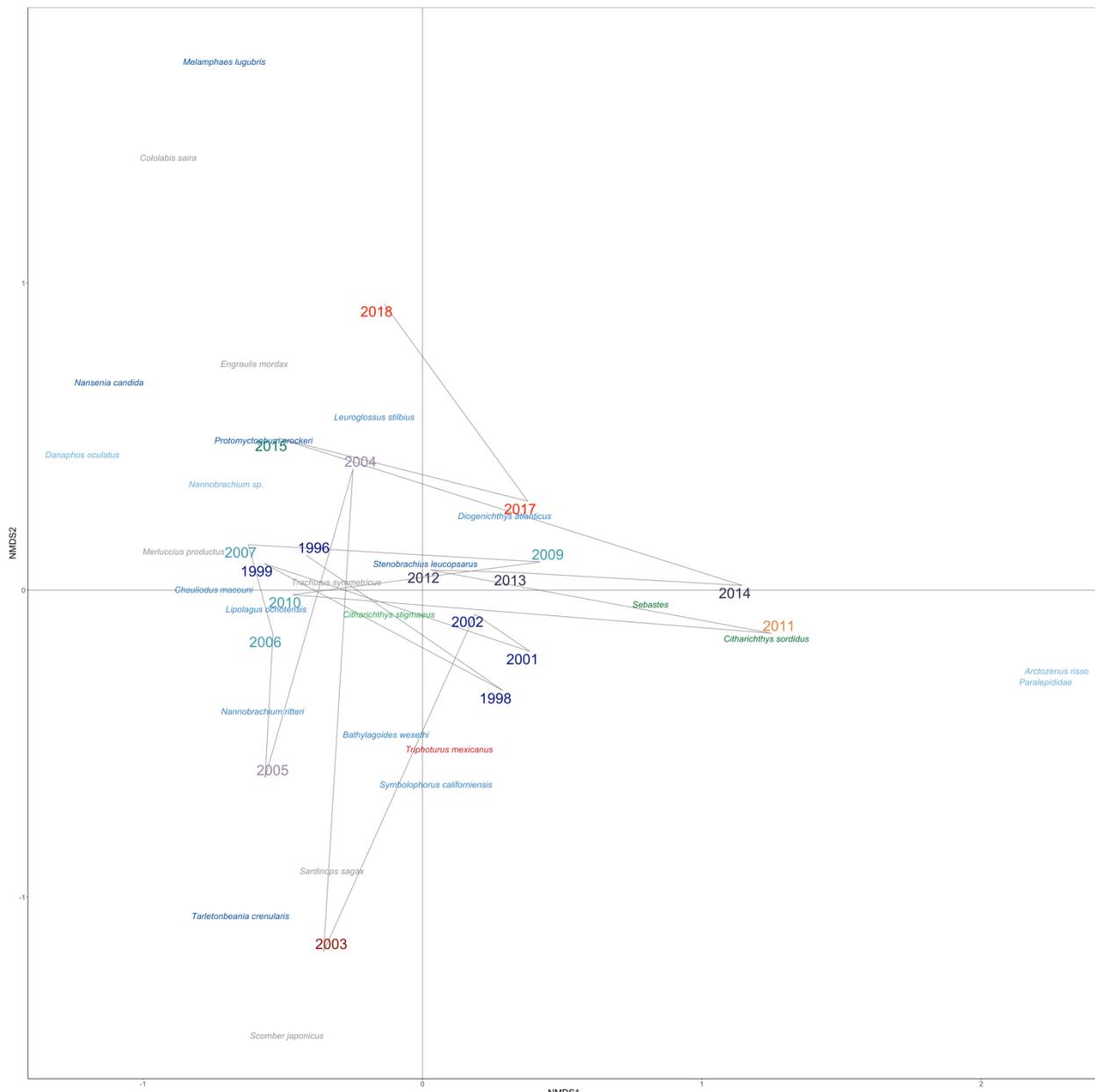
568 NMDS Ordination of Bray-Curtis dissimilarities calculated from abundance of each year.
 569 Years are color coded by chronological clustering ($k = 8$). Species are color coded by
 570 habitat association matching Figure 2.



571

572 **Figure S25. Heat Map of San Nicholas Island Abundances Over Time**

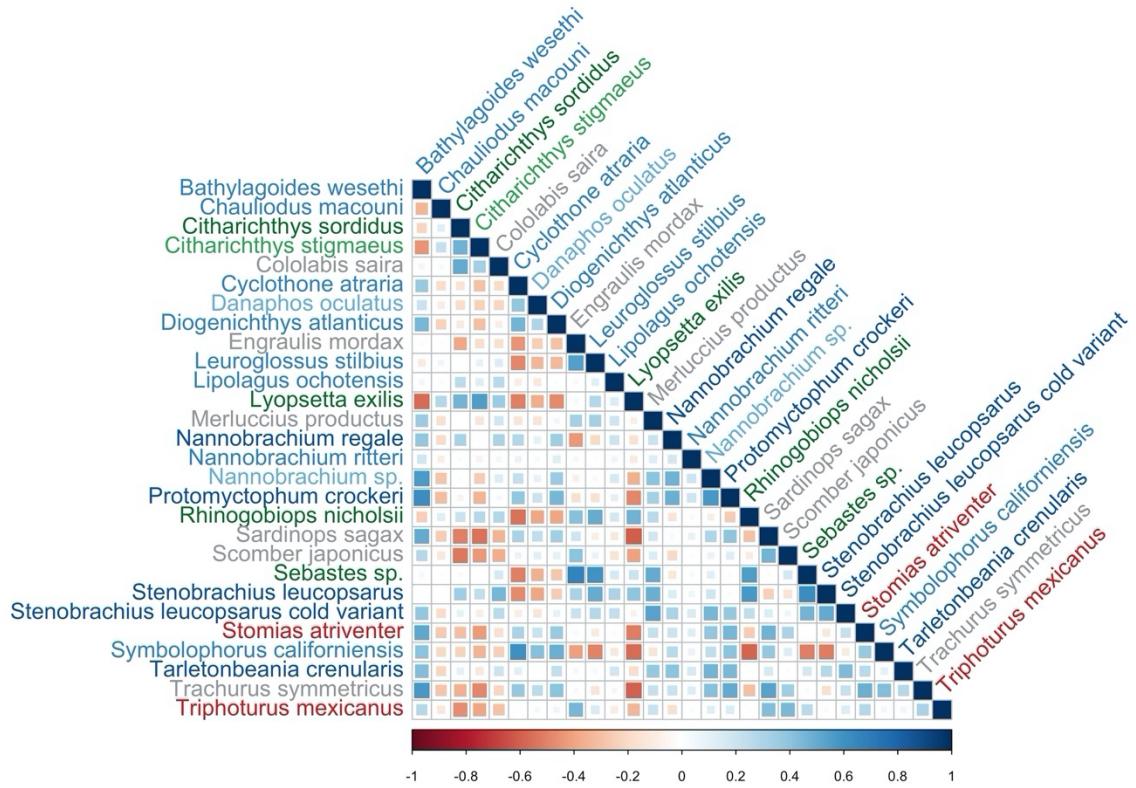
573 Estimated abundance of each year plotted over time. Years are color coded by
 574 chronological clustering. Species are grouped by hierarchical clustering. Lighter colors
 575 indicate higher abundance, white is a lack of detection. Species are color coded by habitat
 576 association matching Figure 1.



577

578 **Figure S26. NMDS Ordination of San Nicholas Island Species and Years**

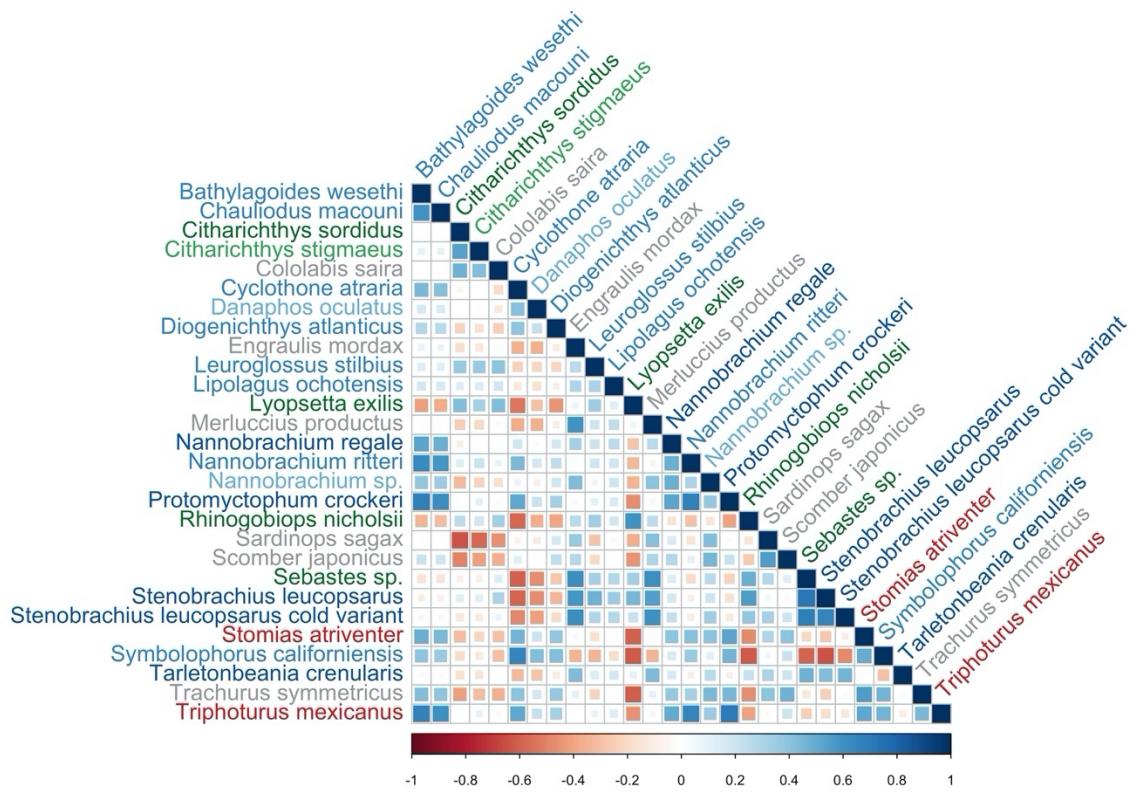
579 NMDS Ordination of Bray-Curtis dissimilarities calculated from abundance of each year.
 580 Years are color coded by chronological clustering ($k = 8$). Species are color coded by
 581 habitat association matching Figure 2.



582

583 **Figure S27. Co-occurrence Patterns of Species Controlling for SST**

584 Benthic fisheries targets like sanddabs were negatively correlated with mesopelagic
 585 species when controlling for SST. These results capture a strong benthic vs. pelagic
 586 community dynamics when controlling for SST. We plot the correlation matrix of the
 587 generalized linear latent variable model predictors across species. Species in blue correlate
 588 are positively correlated with each other while species in red are negatively correlated.
 589 Strength of association is visualized by size of square. Species are color coded by habitat
 590 association matching Figure 2.



591

592

Figure S28. Co-occurrence Patterns of Species

593

Fisheries targets like North Pacific Hake were negatively correlated with mesopelagic

594

species. We plot the correlation matrix of the generalized linear latent variable model

595

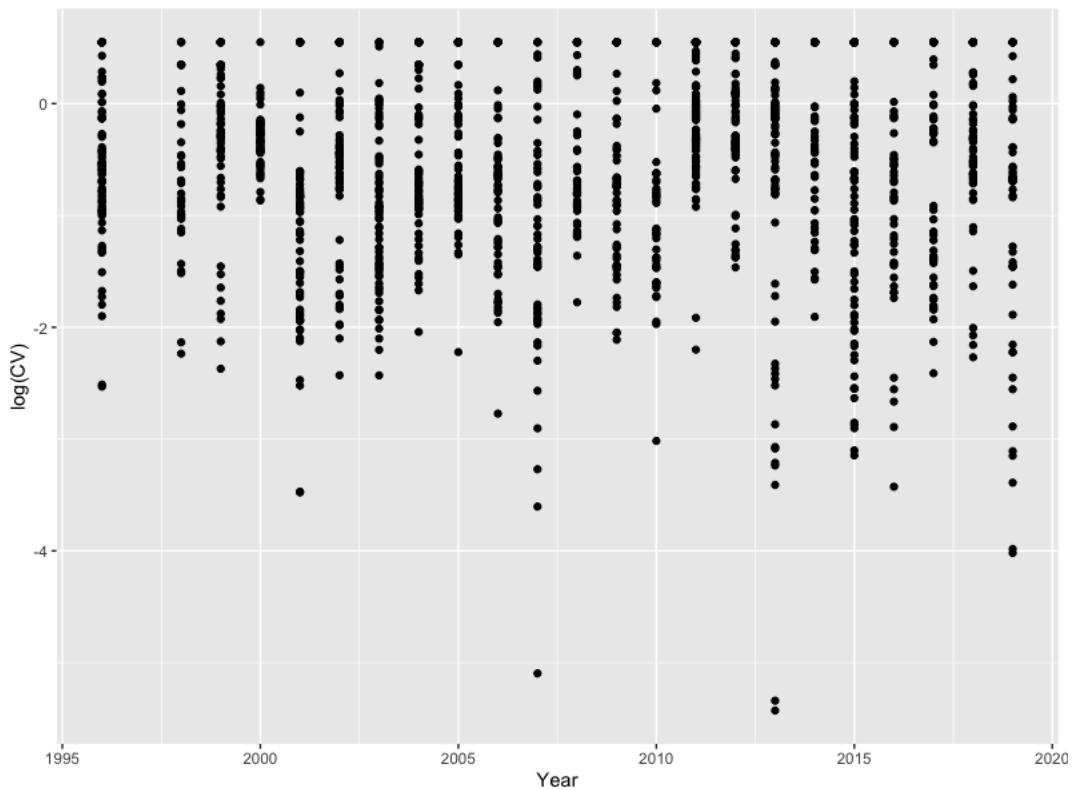
predictors across species. Species in blue correlate are positively correlated with each

596

other while species in red are negatively correlated. Strength of association is visualized

597

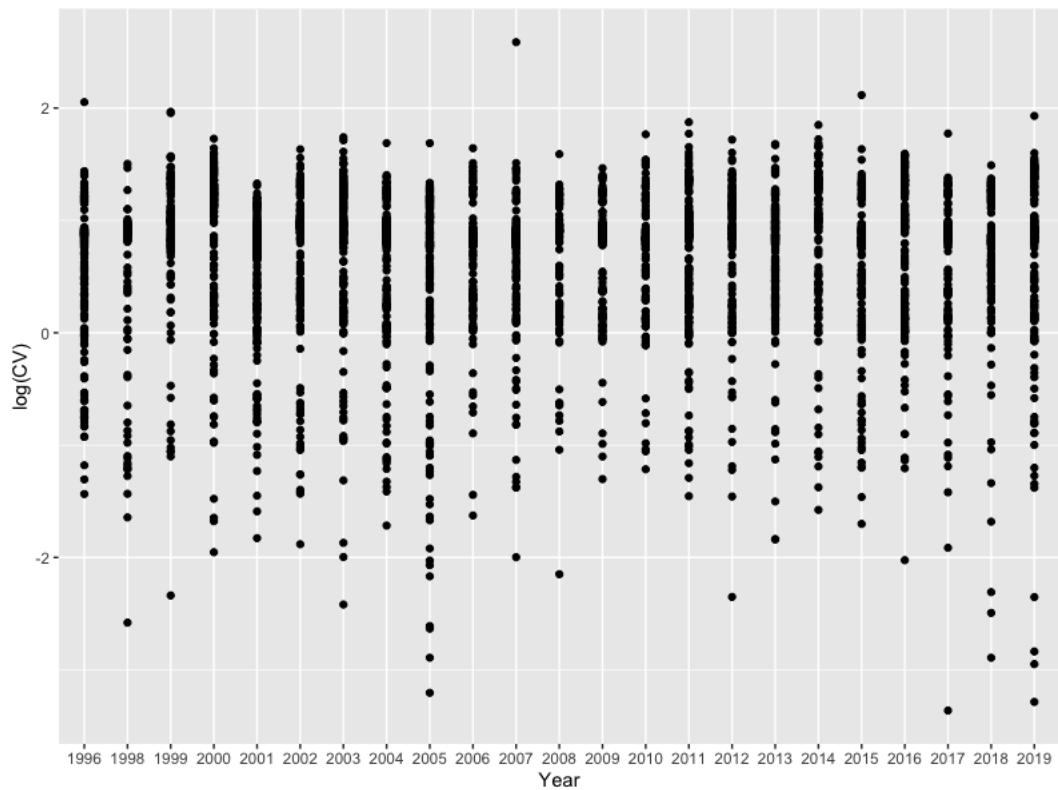
by size of square. Species are color coded by habitat association matching Figure 2.



598

599 **Figure S29. Stable Precision of Amplicon Abundance Over Time**

600 Here we measure the coefficient of variation (CV) of species-specific amplicons across
601 three technical replicates. An increase in CV with the age of the sample would signal
602 degradation; we see no such trend.



603

604

Figure S30. Stable Precision of Abundance Estimates Over Time

605

606

Coefficient of variation of model estimates of abundance over time. We observe no evidence of change in precision over time.

Supplement S2

Gold et al. 2021

Model description of morphology and metabarcoding analysis of CALCOFI samples

Motivating models.

We use metabarcoding approaches outlined in Supplement S1 to produce amplicons from ethanol-preserved ichthyoplankton samples. Briefly, we generated amplicons using the MiFish 12S Universal Teleost primer set (23) on DNA extractions derived using the Qiagen DNeasy Blood and Tissue kit filtered ethanol-preservative. Each amplicon library was sequenced separately on an Illumina NextSeq. See Supplement S1 for full details.

We estimate that, for any species i , the number of sequenced amplicons is proportional to the fraction of DNA from that species in the PCR template (53,55,56). We note that here we use i to represent species, but can be more generalized to represent “molecular targets” or “amplicon sequence variants.” The amplicons produced during a PCR reaction are governed by the efficiency parameter α_i , which is characteristic of the interaction between the particular primer set and each species being amplified. Thus, for any species i , the number of amplicons should be directly related to the efficiency of amplification and the starting concentration of DNA template such that

$$A_i = c_i(1 + \alpha_i)^{N_{PCR}} \quad (1)$$

where A_i is amplicon abundance, c_i is the true number of DNA copies in the reaction attributable to species i , α_i is the amplification efficiency (bounded on $(0, 1)$), and N_{PCR} is a known constant (an integer giving the number of PCR cycles used in the reaction). If we could perfectly observe amplicons, the above equation would be all that we needed to relate the amplicons we observe and the biological value of interest, c_i , the true number of template DNA copies. Unfortunately standard PCR and sequencing technology does not allow for such direct observation. Due to N_{PCR} being a large number and α_i typically being not close to 0, the number of amplicons expected for any species with $c_i > 0$ is very, very large (e.g. with $c_i = 2$, $\alpha_i = 0.75$, and $N_{PCR} = 36$, $A_i = 1.12 \times 10^9$) and there are typically many species being amplified simultaneously, producing 10^{10} or more DNA copies in a single reaction. The actual number driven primarily by the α values among species and N_{PCR} . This model assumes that PCR amplification has not approached saturation and therefore the PCR is still amplifying exponentially. We, and others (53,55,56), argue this assumption is valid because 1) the total concentration of DNA within a filtered ethanol sample is low ($<1\text{ng}/\mu\text{L}$) and 2) the PCR reagents are supplied in excess and therefore are unlikely to be saturating the PCR. However, future models could be developed to account for a saturating PCR curve (68).

DNA sequencing machines do not read all of the copies from such a reaction; they read only a small fraction of the reads (on the order of 10^6 to 10^7 reads) (54,70). This subsampling changes what in eq. 1 appears to be a single-species process – each species being amplified independently – into a multi-species process; the number of amplicons observed for species i will depend upon both the amplicons produced for species $i = 1$ and the amplicons produced for species $i = 2, 3, \dots, I$ in the same reaction. Observations of amplicons are thus compositional data and need to be analyzed as such.

Models for compositional data

We want to retain the data-generating structure from eq.1 as much as possible, so we develop a model for a single sample with many species. As above, let i index species with $i = 1, 2, \dots, I$ and then we can write a deterministic equation for the number of amplicons observed in log-space as

$$\log(A_i) = \log(c_i) + N_{PCR} \log(1 + \alpha_i) + \log(\eta) \quad (2)$$

where the only new term is η which represents the proportion of reads observed from a given sampling run. Note that in this formulation η is a single value shared across all species and serves to scale the number of amplicons observed. Additionally we can rewrite the number of DNA copies in terms of proportional number of counts, $\log(\beta_i) = \log(c_i) - \log(\sum_i c_i)$. Note that the second term in this equation is a sum of the counts across all species, and so is a single shared value for all species. As such it can be absorbed into the value η that scales the overall abundance,

$$\log(A_i) = \log(\beta_i) + N_{PCR} \log(1 + \alpha_i) + \log(\eta) \quad (3)$$

This equation is appealing because it provides a process-oriented description of the biology of interest (the β s), a term for how PCR modifies our observation of the true abundance ($N_{PCR} \log(1 + \alpha_i)$), and a term for how subsampling of DNA reads in the sequencing machine will adjust the number of amplicons observed $\log(\eta)$. This third term also links all of the single-species components to produce a multi-species model. It is important to note that while both eq. 2 and 3 use the term η , the interpretation and plausible range of this parameter are distinct in the two equations. In eq. 2, $0 < \eta \leq 1$, while in eq. 3 η is not constrained to be less than 1 ($\eta > 0$).

In practice, PCR and subsampling are not deterministic but random processes (70). Furthermore, we are rarely interested in results from a single sample but rather multiple samples collected across sites j and times t . We let λ_{ijtk} be the expected number of amplicons observed, with k indexing unique PCR reactions to account for the fact that there may be multiple individual PCR reactions for a single collected sample,

$$\log(\lambda_{ijtk}) = \log(\beta_{ijt}) + N_{PCR} \log(1 + \alpha_i) + \log(\eta_{jtk}) \quad (4)$$

Importantly, α_i is assumed to be constant for each species among all sites, times, and PCR reactions (53,55,56). We incorporate stochasticity by allowing the number of observed amplicons to vary from the deterministic mean by specifying the observed values as following a negative binomial distribution,

$$Y_{ijk} \sim \text{NegativeBinomial}(\lambda_{ijtk}, \phi) \quad (5)$$

$$\phi = \exp[\tau_0 + \tau_1 \log(\lambda_{ijtk})] \quad (6)$$

where the expected value and variance of Y_{ijk} are $E[Y_{ijk}] = \lambda_{ijtk}$, and $Var[Y_{ijk}] = \lambda_{ijtk} + \frac{\lambda_{ijtk}^2}{\phi}$, respectively.

Note that we allow for the scale parameter ϕ to vary with the predicted mean, this allows for the amount of dispersion in the negative binomial to shift with changing predictions. For our datasets, this allows dispersion to be large when λ is small and decrease as λ increases.

By itself, this model has substantial identifiability problems; in the absence of additional information, it is not possible to estimate the β and α parameters. In the next section we discuss how to integrate data from amplicon sequencing as well as other data sources to make the parameters identifiable.

Application to the CalCOFI Dataset

At each CalCOFI station, an oblique bongo-net tow is conducted from 210 m to the surface with the starboard side preserved in buffered formaldehyde and the port side preserved in buffered ethanol-preserved (See Supplemental S1 for more details)(12). Manual counts of ichthyoplankton were quantified using microscopy to identify species abundance from formaldehyde-preserved samples. Metabarcoding was conducted on the ethanol preserved side; consequently, we expect the contents of the paired samples to differ slightly as a function of sampling stochasticity. Counts of larvae/juveniles were done once for each jar.

The manual counts provide extra information that enable us to estimate the confounded parameters from the metabarcoding. Specifically, for each sampled station, we have two sets of observed data: 1) counts of larval/juvenile fishes for each taxon from the formaldehyde jars (Z_{ijt} ; indexes as above) and 2) counts of amplicons for each taxon from ethanol jars (Y_{ijtk}). These observed data arise from a common (but unobserved) biomass for each species at each station-year combination (γ_{ijt} ; a latent (unobserved) variable).

We assume that the data are counts for each species in each sample, Z_{ijt} , derived from the true density of each species γ_{ijt} , the fraction of total specimens counted in each vial, P_{jt} , and the volume of water filtered for that station relative to a standard volume, V_{jt} ; $V_{jt} \approx 1$ for most samples, $V_{jt} < 1$ indicates a smaller volume of water was sampled.

$$Z_{ijt} \sim \text{Poisson}(\theta_{ijt}) \quad (7)$$

$$\log(\theta_{ijt}) = \log(\gamma_{ijt}) + \log(P_{jt}) + \log(V_{jt}) \quad (8)$$

We consider β_{ijt} to be the true proportion of biomass at a given station-year for each taxon i , $\beta_i = \frac{\gamma_{ijt}}{\sum_i \gamma_{ijt}}$.

Joint Model for Counts and Amplicons

To combine our information from the manual counts and metabarcoding, we need to recognize that our observations (Y_{ijtk} and Z_{ijt}) are linked together by a common variable (γ_{ijt}) and thus we can model them jointly (71). We represent the amplification process using equation 5 and 6 above (amplicons were sequenced in triplicate reactions for each jar). The manual count are modeled as in equations 7 and 8.

Our model assumes the fraction of template DNA in each sample is proportional to the counts of individual species' larvae in each paired jar (53,55,56). Moreover, we assume that in each sample there is template DNA from species that are uncounted, unidentifiable, or otherwise unobserved (69,70). In practice, this DNA might derive from stochastic sampling between each side of the bongo net, excreted waste, stray tissue, cells, or microscopic genetic material extracted along with the visible larvae.

Dealing with the fact that not all methods see the same species

The above is sufficient if all of the species identified by morphological counts are identical to the species identified by the genetic methods. But this is often not the case; some larvae are not separable to species based on morphology and some species are not separable to species based on a single genetic primer. Furthermore, some species do not amplify at all in the PCR ($\alpha_i \approx 0$). To accommodate non-overlapping sets of species among sampling methods we introduce a new variable, γ_{Mijt} , which specifies the true (M is for "main") density of species i at site j and time t . We assume that there is a mapping between this main density and the density observed by each sampling method. Specifically, we assume the species in the main list maps uniquely onto no more than one taxonomic group in each observation method, but multiple main species can map onto a single group for each observation method. For example, if the observation of larval counts identified as a specimen as *Sebastes* sp., we assume this may map onto one or more specific taxa (e.g., *Sebastes paucispinis*) in the main list, but conversely, *Sebastes paucispinis* on the main list may not map to more than one entity identified by each observation method.

We can construct a mapping matrix, \mathbf{M}_{MS} , that transforms the species in the main list, γ_M (a vector of length I_M , the number of true species in the sample) into the species grouping observed by sampling method

S , γ_S (a vector of length I_S , the number of groups observed by method S). We drop the j and t subscript because this mapping does not depend on the identity of the community being sampled. Then,

$$\gamma_S = \mathbf{M}_{MS} \gamma_M \quad (9)$$

\mathbf{M}_{MS} is a I_S by I_M matrix.

For example, if there are four species in the true community and method S only observes three groups, the matrix \mathbf{M}_{MS} could look like this

$$\mathbf{M}_{MS} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (10)$$

This might happen if species 2 and species 4 (columns 2 and 4, respectively) were from the same genus and the PCR primer from method S can only resolve those two species at the genus level. To provide a further example, take an invented community of four species with $\gamma_M = \{1, 15, 6, 7\}$ individuals in the community. The true community as observed through method S would be

$$\gamma_S = \mathbf{M}_{MS} \gamma_M \quad (11)$$

$$\begin{bmatrix} 1 \\ 22 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 15 \\ 6 \\ 7 \end{bmatrix} \quad (12)$$

and so γ_S is a linear combination of the true community. Of course there is no requirement that elements of γ_M be integers, but that makes the above example easy and transparent.

It is easy to incorporate this added complexity into the models in the previous section. If we assign method S to be manual counts and W to be the Mifish PCR primer, we need to construct a main list of species to define γ_M and build two mapping matrices, \mathbf{M}_{MS} and \mathbf{M}_{MW} that determine which species or species-groups are observed by each method. We can then add an additional subscript for each additional method and use the same form as above. For example,

$$\log(\theta_{Sijt}) = \log(\gamma_{Sijt}) + \log(P_{Sjt}) + \log(V_{Sjt}) \quad (13)$$

$$\log(\lambda_{Wijtk}) = \log(\beta_{Wijt}) + N_{W,PCR} \log(1 + \alpha_{Wi}) + \eta_{Wjtk} \quad (14)$$

And with additional sampling methods, we can make different mappings from the true abundance to the observations of each method.

Estimation and Identifiability

We developed and fit the above model in a Bayesian framework using the Stan language, as implemented in *RStan* (75, 76). All code is available as supplementary material. Table S2.1 provide prior distributions used in the model.

We ran five MCMC chains with 1,000 warmup and 4,000 sampling iterations. We retained every other MCMC sample. We initiated each chain at randomly determined starting values. The model converged ($\hat{R} < 1.01$; Gelman-Rubin diagnostics) and had no divergent transitions. We performed standard posterior predictive checks to assess model fit.

Table S2.1: Prior and parameter descriptions for the Stan Model.

Parameter & Prior	Description
$\alpha_i \sim Beta(1, 1)$	Amplification efficiency for species i
$\log(\gamma_{Mijt}) \sim Normal(0, 4)$	True biomass of each species at each site-year
$\log(\eta_{jtk}) \sim Normal(-4, 4)$	Estimated offset for each PCR reaction at each site-year
$\tau_0 \sim Normal(0, 2)$	Negative Binomial shape parameter intercept
$\tau_1 \sim Normal(0, 2)$	Negative Binomial shape parameter slope