Monterey Bay Aquarium
Research Institute

Nucleotides to Rorquals: Delineating humpback whale distribution with environmental DNA

Olivia S. Boeberitz

Mentor: Dr. Francisco Chavez

Summer 2020

Keywords: Environmental DNA (eDNA), cetacean, temporal distribution, metabarcoding

ABSTRACT

The assessment of species distribution is integral to understanding biodiversity in aquatic systems but remains difficult for rare or cryptic species such as whales. Environmental DNA (eDNA) has become a powerful non-invasive approach for genetic monitoring of species presence in marine ecosystems, providing further insight towards the ecological and spatial processes of difficult-to-sample taxa. This study examines eDNA metabarcoding data amplified with universal MiFish primers that target the mitochondrial 12S rRNA gene to characterize cetacean communities in Monterey Bay, California. Water samples were collected monthly from a nearshore location (C1) during the years of 2008-19. Cetacean communities detected with eDNA were analyzed to determine species abundance and distribution to show community structural differences over time. *Megaptera novaeangilae* (humpback whale) showed to be the most prominent in all datasets, with 8 different ASVs identified, encompassing 30,816 high quality Illumina sequence reads. The presence of humpback ASVs was found to be correlated with the presence of anchovies and upwelling status in the system, as well as following known fall feeding migration trends. Results suggest accurate portrayal of humpback presence in

Monterey using eDNA, with coastal detections increasing in the fall season, as well as after 2015 when upwelling starts to slow. This study demonstrates our ability to detect and examine trends of marine megafauna that are important to the Monterey Bay ecosystem.

INTRODUCTION

As marine mammals face increasing pressure from ecological and anthropogenic stressors such as climate change, seismic surveys, oil and gas development, vessel traffic, and fishing activity (Hunt et al., 2014), the assessment of new tools to monitor population dynamics has become critical. Monitoring population structure and distribution of marine mammals not only provides further information regarding their health but is vital to understand conservation status and develop strategies to alleviate disturbance (Parsons et al., 2018). This is especially relevant for cetaceans (whales, dolphins and porpoises) present in coastal waters that are directly impacted by anthropogenic implications, like entanglements in fishing gear. (Zimmer, 2019). Such population dynamics are challenging to assess in the case of cetaceans, particularly mysticete whales (baleen whales), in which traditional survey methods can be costly, as well as time and labor intensive (Hunt et al., 2013). Population genetic research requires extracting samples from marine organisms, often in the form of blood and tissue samples taken with biopsy darts. However, sampling tissue may have negative effects, and can be very difficult for large, elusive species such as cetaceans (Adams et al., 2019). In aquatic matrices, fecal plumes, respiratory blow, and sloughed cells have been proven useful sources (Pierszalowski et al., 2013) capable of providing evidence of the species' presence in the environment during the period that the DNA persists (Foote et al., 2012). This form of genetic data can be utilized to further understand past and present distribution status, as well as future prospects of species (Drummond et al., 2005). The promise of new non-invasive sampling technologies, such as environmental DNA (eDNA), provides an abundance of options to gain insight towards population structure and geographical breaks in genetic diversity (Parsons et al., 2018).

eDNA encompasses any genetic material that has been shed from organisms and retained in the environment. eDNA can take many forms, originating from sources such as feces, mucus, blood, sloughed cells, tissue, or scales (Andruszkiewicz et al., 2017).

Applications of eDNA historically have been used for AIS detection, focusing on single-species DNA barcoding (Sard et al., 2019). However, recent developments have expanded eDNA research and quantitative PCR analysis to encompass multi-species sequence analyses of eDNA, known as metabarcoding. Metabarcoding allows for community composition and diversity to be more accurately estimated alongside species relative abundance (Valentini et al., 2015). This advancement provides insight towards ecosystem-level biodiversity in aquatic matrices by filtering or precipitating water samples. eDNA captured through filtering can be PCR-amplified using either species-specific or universal primers. This study analyzes eDNA samples amplified with universal MiFish primers, dedicated to detecting bony fishes, that target the mitochondrial 12S rRNA gene, sequenced using next generation sequencing (NGS). To develop universal primers, whole genomes of groups of species (e.g., marine vertebrates) are aligned and primers are optimized to target a short gene region that is evolutionarily preserved among all species in a group, but varies enough among species to appropriately identify taxa to the genus or species level (Miya et al., 2015). Therefore, by comparing amplicons to a reference genome, an abundance of taxa can be identified from an environmental water sample (Deiener et al., 2017).

While fish are the most common group studied using eDNA techniques, the 12S primers employed in this study are not specific to fish but amplify members of all vertebrate taxa. While acoustic telemetry and photo-identification techniques are the most popular methodologies used for population assessments of whales (Silva et al., 2000), there has been less attention paid to utilizing eDNA as a supplementary tool. Environmental DNA has great potential as a non-invasive sampling technique to monitor cetacean communities.

This study takes place in the California Current System (CCS), at the Monterey Bay National Marine Sanctuary (MBNMS). Two of the fifteen distinct population segments of humpback whales utilize this region for essential feeding and migratory habitat (Bettridge et al., 2015). Humpbacks that feed off the western United States are known to migrate to breeding grounds off the coast of southern Baja, mainland Mexico and Central America (Ryan et al., 2019). The high levels of primary productivity resulting from wind-driven upwelling of nutrient-rich water into the euphoric zone make this habitat favorable for foraging (Huyer, 1983). While the highest occurrence of krill

hotspots in the CCS are associated with canyons between Monterey Bay and Bodega Bay (Santora et al., 2012), anchovy and sardines become more abundant closer to the Monterey Bay coast (Santora et al., 2011). Complementary visual sighting data shows the highest densities of humpback whales off central California (Ryan et al., 2019). The overall aim of this study is to test the use of eDNA for detecting the presence of marine mammals, using humpback whales as a model, in a biologically rich habitat that supports dense populations of whale species.

MATERIALS AND METHODS

SAMPLE COLLECTION

Niskin bottles on a CTD rosette were used to collect environmental DNA seawater samples. At each sampling location, a 1-liter water sample was filtered onto a 0.22 μm polyvinylidene difluoride (PVDF) membrane filter (Millipore, USA). Until further analysis, all filters were flash frozen in liquid nitrogen and preserved at -80°C. Additionally, a wave glider was used to collect SST data along the transect during the cruise. Samples collected from the C1 sampling station were consistently sampled on a monthly basis from 2008 to present day.

DNA EXTRACTION AND AMPLIFICATION

DNA from filters was extracted using the DNeasy® Blood and Tissue kit, carrying out standard protocol (Walz et al. 2019). Concentrations of DNA extracted were quantified using NanoDrop 1000 spectrophotometer (ThermoFisher Scientific, Waltham, MA) measurements. These samples were then metabarcoded for the 12S rRNA gene.

PCR reactions were run with Fluidigm two-step amplification protocol for each sample. Primary PCR reactions were performed in triplicate 25 μl reactions using 1 μl DNA extract ,12.5 μl AmpliTaq Gold Fast PCR master mix (Applied Biosystems, USA), 1 μl each of forward and reverse primers (5 μM), and 9.5 μl molecular-biology grade water. To check for contamination during the PCR process, three PCR blanks (no-template controls) were also run. Specifications of the thermal cycling parameters can be found in Closek et al. 2018. After PCR, technical replicates were run through an agarose gel to visualize the presence of target bands and clean no-template controls (NTCs). The absence of any non-specific amplification and bands in the NTCs confirmed successful

amplification. PCR products were purified, and size selected using the Agencourt AMPure XP bead system (Beckman Coulter, USA). A second agarose gel was run to confirm primer removal and retention of target amplicons after purification. Purified products were then quantified using a Qubit dsDNA HS kit (COI and 12S rRNA gene).

DNA SEQUENCING

The Qubit dsDNA HS kit was used to construct and quantify equimolar pools, confirming pool concentration prior to library preparation. A single library was constructed for each sampled location using the KAPA HyperPrep and Library Quantification kits following manufacturer's protocol. On a standard MiSeq v2 flow cell, the pooled product was loaded for each genetic locus and sequenced in a 2x250bp paired end format using a v2 500-cycle MiSeq reagent cartridge. A PhiX addition of 20% was used to carry out the MiSeq run, and then custom sequencing primers were added to the appropriate wells of the reagent cartridge. Illumina Real Time Analysis (RTA) v1.18.54 was used for base calling, and the output of RTA was demultiplexed and converted to the aligning format (FastQ) with Illumina Bcl2fastq v2.18.0.

BIOINFORMATICS

The resulting sequences were run through a modified version of the banzai pipeline, a custom shell script incorporating different programs to process raw DNA sequences from an Illumina sequence run (O'Donnell et al., 2016). The program Atropos was used to remove the primer sequences within this script from fastq files, which were then fed into the Dada2 program. Here, low-quality regions of reads were trimmed and filtered by quality score, sequencing errors were removed, and the reads were then merged. Blastn searches to NCBI GenBank's nucleotide database were used to assign taxonomy to the resulting ASV sequences. Only annotations with >80% identities were kept and interpreted through MEGAN6. The most recent common ancestors of these hits were determined.

DATA PREPARATION

Following decontamination of sequencing data, in which index-jumping was accounted for, all unwanted taxa were removed. Any ASVs annotated to terrestrial

species were removed, including humans, pigs, cows, dogs, etc. All OTUs were merged together on the basis of unique taxonomic assignment to create ASV tables, taxonomy tables, and a metadata file. By subletting for the Order Cetacea and Genus Megaptera, sample sums and detections were displayed for the desired species.

STATISTICAL ANALYSES

Correlation tests were run on R studio using the package ggpubr in order to determine the relationship between the proportion of humpback reads and the proportion of anchovy reads across a ten-year time series. The Pearson, Spearman, and Kendall correlation tests were all run to compare p-values and correlation coefficients when using and not using ties.

Furthermore, to determine if humpback ASVs varied genetically or if they were a product of contamination, sequences were aligned in Geneious prime to determine base pair differences.

RESULTS

OBSERVED RICHNESS

Sequencing resulted in 90,906 total cetacean reads for the 12S gene. DNA metabarcoding detected 50 unique taxonomic assignments/annotations using the 12S marker, corresponding to 6 families, 12 genera, and 14 species. Cetaceans were detected 791 times, with 302 of these detections assigned to Balaenopteridae (baleen whales) and 336 to the Delphinidae family. Other families detected include Phocoenidae with 58 detections, Ziphiidae with 23, Eschrichtiidae with 1, Physeteridae with 1, and 9 unassigned. Of the 302 baleen detections, 237 were assigned to *Megaptera novaeangliae* (humpback whale). These detections incorporated 30,816 humpback reads.

8 unique humpback ASVs were determined from sequenced data. At both C1 and MARS sampling locations, 5 of these were present. ASV_86 and ASV_114 were the only two ASVs present at every location. These ASVs also contain the largest amounts of reads and detections. ASV_86 with 13,130 reads and 109 detections, while ASV_114 has 16,374 reads and 114 detections. By aligning ASV_86 and ASV_114 in Geneious Prime, it was determined that they vary by two base pairs.

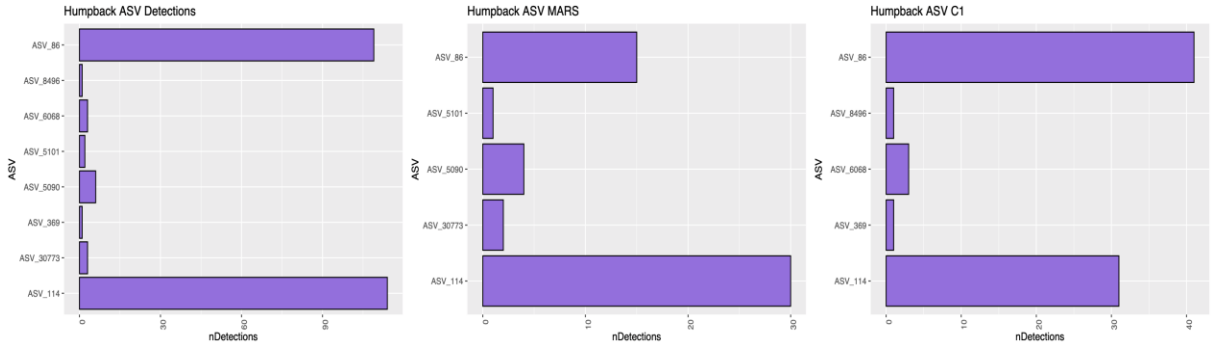

Figure 1: Number of detections for unique humpback whale ASVs in all samples, at the MARS off-shore sampling station and at the coastal C1 sampling station.

EFFECT OF SAMPLING EFFORT

In order to correctly identify the presence and distribution for humpbacks, sampling effort first had to be considered. Across all locations, there was an increase in number of samples taken starting in 2017 as sampling effort surged. Therefore, the number of detections displayed across all locations does not account for sampling size, which is inconsistent. The proportion of humpback whales in all samples can be estimated to avoid sampling bias when considering all locations.

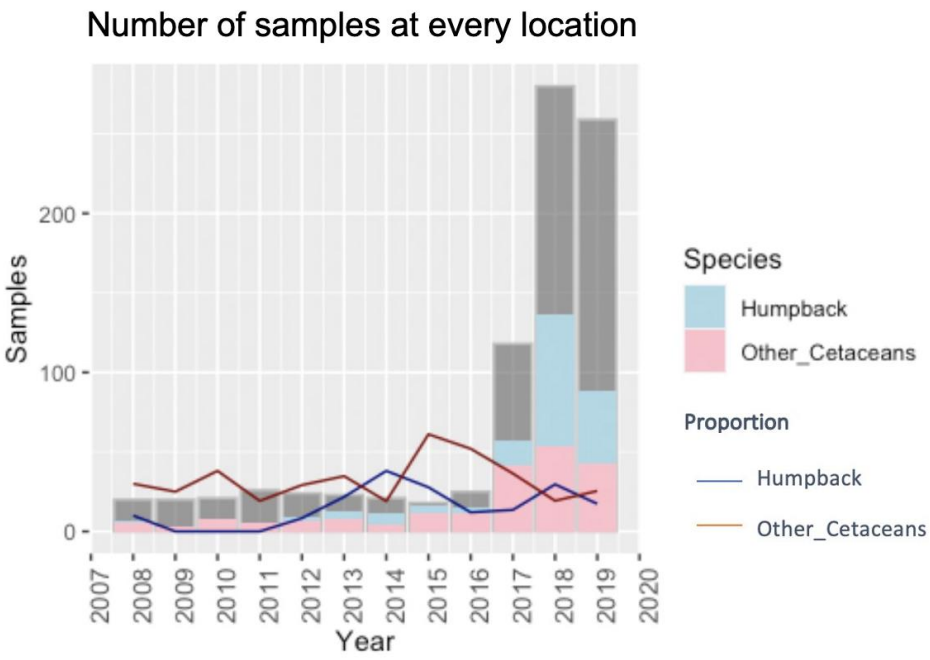

Figure 2: The total number of samples taken per year, represented in grey, with the number of samples in which humpbacks and other cetacean species were detected. Trendlines demonstrate the proportion of all samples in which these species were detected for each year.

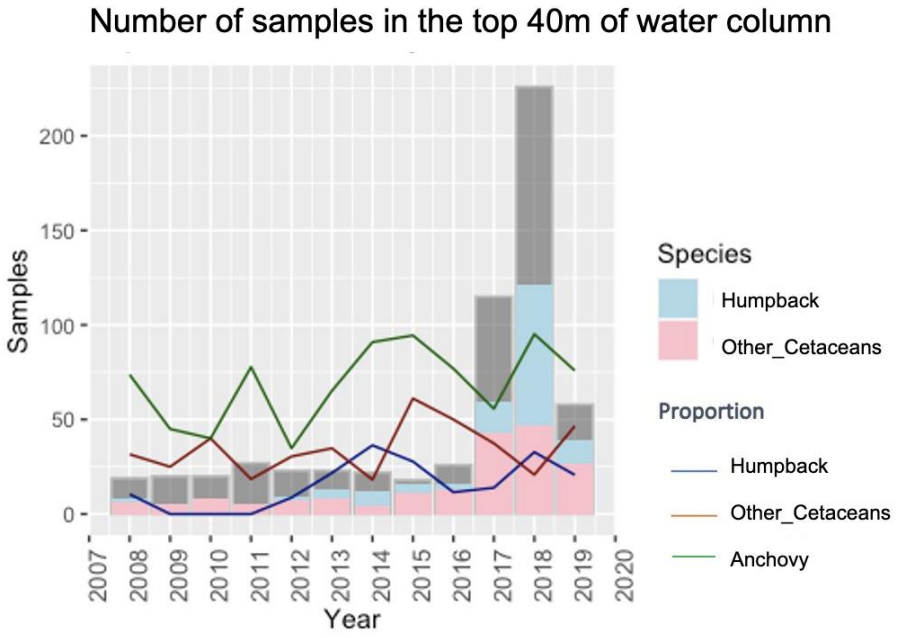

Figure 3: The total number of samples taken per year in the top 40 meters of the water column, represented in grey, with the number of samples in which humpbacks and other cetacean species were detected. Trendlines demonstrate the proportion of all samples in which these species, including anchovies (>100 reads), were detected for each year.

TEMPORAL DISTRIBUTION

By selecting only for samples recorded at the coastal C1 location, the effect of sampling effort was avoided, for C1 was sampled on a consistent monthly basis over the past ten years. This decade long time series allowed humpbacks to be traced back in time. During this ten-year time series, humpback whale reads per ASV were at their highest during the late summer to fall season in each year detections were found. Although, in the years 2008 to 2014, zero humpbacks were detected, or less than half of what we detected in later years. As well as this, we see an inconsistent trend in 2016, which is much lower than its counterparts. By grouping the time series by year, a very clear buildup of reads to 2015 can be seen, followed by a large decrease in 2016, and then heightened again in the following years.

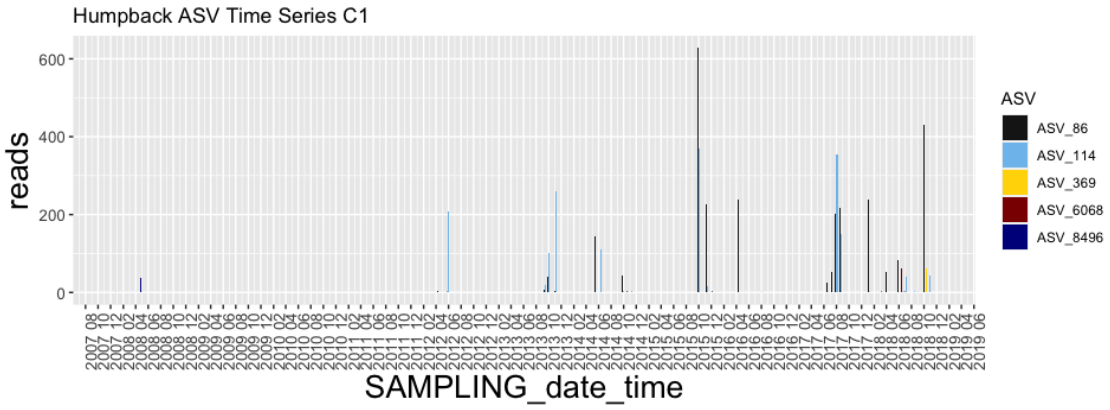

Figure 4: Distribution of individual humpback ASVs at the C1 sampling station over a 10-year time series, with the abundance of reads depicting the intensity of presence.

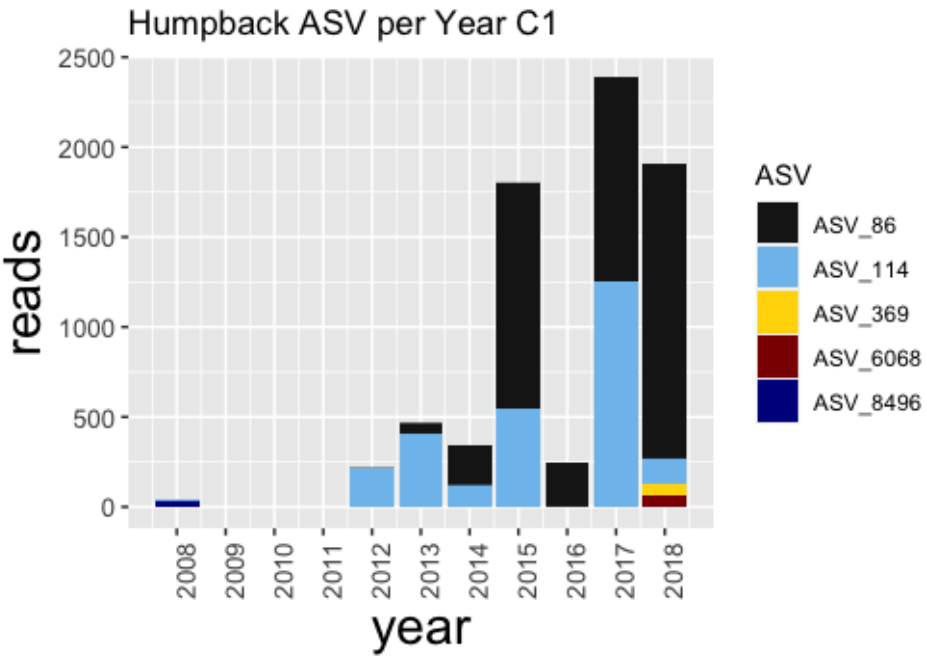

Figure 5: Distribution of individual humpback ASVs at the C1 sampling station per year, with the abundance of reads depicting the intensity of presence.

DISCUSSION

EFFECTIVENESS OF EDNA METABARCODING

eDNA surveys were able to detect a wide array of cetacean species in the Monterey Bay, in which the humpback whale was the most prominent in all locations. More specifically, the universal MiFish primers, dedicated to detecting bony fishes, were able to successfully display trends of humpback whales. The MiFish primers were even able to delineate resolution down to genetically different populations of humpbacks in the area. Although the MiFish primers were able to provide additional insight into cetacean distribution, there was a high abundance of unassigned reads.

DISTRIBUTION IN RELATION TO AN OCEANOGRAPHIC FRONT

eDNA trends developed using metabarcoding, targeting the mitochondrial 12S rRNA gene region are supported by a multitude of fisheries and oceanographic data. In 2016, the decrease in number of reads is mirrored by known data that shows a low abundance of anchovies in this area during the same time period (Andruszkiewicz, 2017).

Prior to 2015, we see a decrease in amount of reads for the humpback whales as upwelling starts to slow down, causing mean sea surface temperature to rise. These warm waters start to show up as the cold, nutrient rich waters are no longer being upwelling in 2015. Therefore, there is a collapse in coastal productivity (Andruszkiewicz, 2017). The productive areas that are needed to support humpback foraging start to shift closer to the shore, bringing humpbacks in towards our C1 location during these later years.

These trends also support entanglement data, in which after 2015, we see an increase in humpback entanglements with crab traps (Zimmer, 2019). Therefore, eDNA can further inform temporally and spatially intensive datasets to further augment known trends and distribution.

CONCLUSIONS/RECOMMENDATIONS

Environmental DNA metabarcoding is a non-invasive means of sampling, which works effectively as a tool for monitoring elusive marine mammals. The universal MiFish

primer set was able to accurately detect the presence and distribution of humpback whales in the Monterey Bay area, to the point where we were able to delineate among genetically different populations. Future work could enable the identification of these specific groups of humpback whales.

eDNA has great potential to further augment traditional methods, providing insight into past and present biodiversity and community structure. Although using MiFish primers may be limiting the scope of our data. Therefore, in the future, using recently developed marine mammal primers could heighten our ability to track cetaceans and fill in our unassigned blanks. More so than this, by combining eDNA with other non-invasive techniques like acoustics, we may be able to determine more humpback population mixing in feeding grounds than once previously thought.

ACKNOWLEDGEMENTS

I would like to thank my mentor, Francisco Chavez, for his guidance throughout this project. Thank you to Markus Min for his data analysis R scripts, as well as Katie Pitz, Charles Nye, Reiko Michisaki, and Nathan Truelove for not only providing feedback on my project, but for offering their support and advice for data management. I would also like to thank George Matsumoto and Megan Bassett for organizing a wonderful and inclusive intern program. Lastly, thank you to the Dean and Helen Witter Family Fund, the Rentschler Family Fund, and the David and Lucile Packard Foundation for funding.

References:

Adams, C.I.M., Knapp, M., Gemmell, N.J., Jeunen, G., Bunce, M., Lamare, M.D., Taylor, H.R., (2019). Beyond biodiversiry: Can environmental DNA (eDNA) cut it as a population genetics tool? *Genes*, 10: 192. https://doi.org/10.3390/genes10030192

Andruszkiewicz, E.A., Starks, H.A., Chavez, F.P., Sassoubre, L.M., Block, B.A., Boehm, A.B., (2017). Biomonitoring of marine vertebrates in Monterey Bay using

eDNA metabarcoding. *PLoS ONE*, 12: e0176343. https://doi.org/10.1371/journal.pone.0176343

Calambokidis, J., Steiger, G. H., Rasmussen, K., Urbán, R., Balcomb, K. C., Ladrón de Gueveara, P., Salinas, M., Jacobsen, J. K., Baker, C. S., Herman, L. M., Cerchio, S., Darling, J. D., (2000). Migratory destinations of humpback whales from the California, Oregon and Washington feeding ground. *Marine Ecology Progress Series*, 192: 295-304.

Closek, C., Djurhuus, A., Pitz, K. Kelly, R., Michisaki, R., Walz, K., Starks, H., Chavez, F., Boehm, A., Breitbart, M., (2018). Environmental DNA (eDNA) 18S metabarcoding Illumina MiSeq NGS PCR Protocol. *protocols.io* dx.doi.org/10.17504/protocols.io.n2vdge6

Da Silvia, C.Q., Zeh, J., Madigan, D., Laake, J., Rugh, D., Baraff, L., Koski, W., Miller, G., (2000). Capture-recapture estimation of bowhead whale population size using photo-identification. *Journal of Cetacean Research and Management*, 2: 25-61.

Deiner, K., Bik, H.M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D.M., De Vere, N., Pfrender, M.E., (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, 26: .5872-5895. https://doi.org/10.1111/mec.14350

Fleming, A.H., Clark, C.T., Calambokidis, J., Barlow, J., (2015). Humpback whale diets respond to variance in ocean climate and ecosystem conditions in the California Current. *Global Change Biology*, 22: 3. https://doi.org/10.1111/gcb.13171

Foote, A.D., Thomsen, P.F., Sveegaard, S., Wahlberg, M., Kielgast, J., Kyhn, L.A., Salling, A.B., Galatius, A., Orlando, L., Gilbert, T.P., (2012). Investigating the potential use of environmental DNA (eDNA) for genetic monitoring of marine mammals. *PLoS ONE*, 7: e41781. https://doi.org/10.1371/journal.pone.0041781

Hunt, K.E., Moore, M.J., Rolland, R.M., Kellar, N.M., Hall, A.J., Kershaw, J., Raverty, S.A., Davis, C.E., Yeates, L.C., Fauquier, D.A., Rowles, T.K., Kraus, S.D., (2013). Overcoming the challenges of studying conservation physiology in

large whales: a review of available methods. *Conservation Physiology*, 1: cot006. https://doi.org/10.1093/conphys/cot006

Hunt, K.E., Stimmelmayr, R., George, C., Hanns, C., Suydam, R., Brower, H., (2014). Baleen hormones: a novel tool for retrospective assessment of stress and reproduction in bowhead whales (*Balaena mysticetus)*. *Conservation Physiology*, 2: cout030. https://doi.org/10.1093/conphys/cou030

Huyer, A., (1983). Coastal upwelling in the California current system. *Progress in Oceanography*, 12: 259-284. https://doi.org/10.1016/0079-6611(83)90010-1

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J.Y., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka, H., Araki, H., Kondoh, M., Iwasaki, W., (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *Royal Society Open Science*, 2:150088. doi:10.1098/rsos.150088

O'Donnell, J.L., Kelly, R.P., Lowell, N.C. Port, J.A., (2016). Indexed PCR primers induce template-specific bias in large-scale DNA sequencing studies. *PloS one*, 11(3). p.e0148698

Parsons, K.M., Everett, M., Dahlheim, M., Park, L., (2018). Water, water everywhere: environmental DNA can unlock population structure in elusive marine species. *Royal Society Open Science*, 5: 180537. https://doi.org/10.1098/rsos.180537

Ryan, J.P., Cline, D.E., Joseph, J.E., Margolina, T., Santora, J.A., Kudela, R.M., Chavez, F.P., Pennington, J.T., Wahl, C., Michisaki, R., Benoit-Bird, K., Forney, K.A., Stimpert, A.K., DeVogelaere, A., Black, N., Fischer, M., (2019). Humpback whale song occurrence reflects ecosystem variability in feeding and migratory habitat of the northeast Pacific. *PLoS ONE* 14: e0222456. https://doi.org/10.1371/journal.pone.0222456

Sard, N.M., Herbst, S.J., Nathan, L., Uhrig, G., Kanefsky, J., Robinson, J.D., Scribner, K.T., (2019). Comparison of fish detections, community diversity, and relative abundance using environmental DNA metabarcoding and

traditional gears. *Environmental DNA*, 1: 368-384. https://doi.org/10.1002/edn3.38

Santora, J.A., Sydeman, W.J., Schroeder, I.D., Wells, B.K., Field, J.C., (2011). Mesoscale structure and oceanographic determinants of krill hotspots in the California Current: Implications for trophic transfer and conservation. *Progress in Oceanography*, 91: 297-409. https://doi.org/10.1016/j.pocean.2011.04.002

Thomsen, P.F., Willerslev, E., (2015). Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, 183: 4-18. https://doi.org/10.1016/j.biocon.2014.11.019

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P.F., Bellemain, E., Besnard, A., Coissac, E., Boyer, F., Gaboriaud, C., Jean, P., Pulet, N., Roset, N. Copp, G.H., Geniez, P., Pont, D., Argillier, C., Baudoin, J., Peroux, T., Crivelli, A.J., Olivier, A., Acqueberge, M., Brun, M.L., Moller, P.R., Willerslev, E., Dejean, T., (2015). Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, 25: 929-942. doi:10.1111/mec.13428

Walz, K., Yamahara, K., Michisaki, R., Chavez, F., (2019). MBARI Environmental DNA (eDNA) extraction using Qiagen DNeasy Blood and Tissue Kit. *protocols.io* dx.doi.org/10.17504/protocols.io.xjufknw

Zimmer, K., (2019). California crab fishermen are testing "ropeless" gear to save whales – and themselves. *The Counter*. https://thecounter.org/humpback-whale-entanglement-california-dungeness-crab-fishery-ropeless-gear/