

# FishCARD: Fish 12S California Current Specific Reference Database for Enhanced Metabarcoding Efforts

Zachary Gold<sup>1</sup>, Emma Choi<sup>2</sup>, Dovi Kacev<sup>2</sup>, Benjamin Frable<sup>2</sup>, Ronald Burton<sup>2</sup>, Kelly Goodwin<sup>3</sup>, Andrew Thompson<sup>4</sup>, and Paul Barber<sup>1</sup>

<sup>1</sup>UCLA

<sup>2</sup>University of California San Diego Scripps Institution of Oceanography

<sup>3</sup>Atlantic Oceanographic and Meteorological Laboratory

<sup>4</sup>Southwest Fisheries Science Center

June 5, 2020

## Abstract

DNA metabarcoding is an important tool for molecular ecology. However, metabarcoding effectiveness hinges on the quality of reference databases for taxa and loci of interest. This limitation is true for metabarcoding of marine fishes in the California Current Large Marine Ecosystem where there is a paucity of reference 12S barcodes. Here we present FishCARD, a California Current-specific fish 12S-specific reference barcode database. We barcoded 612 species using the MiFish metabarcoding primers; an addition of 258 species to the 459 California Current fish species with existing 12S barcodes from GenBank. The resulting FishCARD database covers 82.7% of California Current fishes, and it includes virtually all fishes sampled by large marine monitoring programs such as the Partnership for Interdisciplinary Studies of Coastal Oceans and California Cooperative Oceanic Fisheries Investigation. To demonstrate the importance of complete reference databases for eDNA metabarcoding, we compared species and reads identified from three 1L seawater samples collected off Santa Cruz Island, CA using GenBank sequences with and without our generated barcodes, as well as the FishCARD database curated here. The inclusion of our generated barcodes allowed the additional identification of 15 native taxa and 21.8% of total reads from eDNA samples. However, we found that half of all amplicon sequence variants (ASVs) generated by MiFish 12S primers were of non-vertebrate 16S origin, demonstrating a clear limitation of a widely employed fish metabarcoding primers. Despite these limitations, FishCARD provides an important genetic resource to enhance the effectiveness of marine metabarcoding efforts in the California Current Large Marine Ecosystem.

## Introduction

Next generation DNA sequencing provides advanced tools for marine ecology and ecosystem monitoring (Closek et al., 2019; Kelly, Port, Yamahara, Martone, et al., 2014; Yamahara et al., 2019). The ability to sequence tens to hundreds of millions of reads in a single sequencing run allows for the development of novel genomic applications to a suite of research questions including species mapping, biomonitoring, gut content analyses, and population genomics, all of which aid our understanding of the ecology of marine ecosystems (Baetscher et al., 2019; Guo, 2017; Sanders et al., 2015; Thompson, Chen, Guo, Hyde, & Watson, 2017).

Key to these advances is next-generation sequencing metabarcoding. Metabarcoding is a process in which multiple species are identified from bulk DNA (e.g., homogenized gut contents or settlement tile scrapings) or environmental DNA (eDNA) samples (e.g., water and soil) typically by PCR amplification and sequencing of a target gene, and then comparing the resulting DNA sequences to a database of known reference sequences (Taberlet, Coissac, Hajibabaei, & Rieseberg, 2012). In particular, the application of eDNA metabarcoding allows researchers to detect a broad range of marine diversity from a single liter of seawater and has the

potential to dramatically improve marine biomonitoring efforts (Kelly, Port, Yamahara, Martone, et al., 2014).

The success of metabarcoding approaches relies on the quality of reference databases, specifically their completeness and accuracy (Boyer et al., 2016; Machida, Leray, Ho, & Knowlton, 2017). The absence of reference barcodes for a given species for a target locus makes it impossible to accurately classify all sequences generated through metabarcoding with current bioinformatic technology (Deiner et al., 2017). Inadequate reference databases are an acute problem for barcoding, metabarcoding, and eDNA studies that limit the accuracy of taxonomic identification and have the potential to bias the interpretation of results (Andruszkiewicz et al., 2017; Djurhuus et al., 2020; Klymus, Marshall, & Stepien, 2017). Thus building complete and accurate reference databases is paramount to the success of molecular ecology monitoring efforts (Schenekar, Schletterer, Lecaudey, & Weiss, 2020). To address the need for accurate and complete reference databases, previous efforts were made to barcode California Current Large Marine Ecosystem fishes focused on the mitochondrial Cytochrome Oxidase I (*COI*) locus (Ardura, Planes, & Garcia-Vazquez, 2013; Elena M Duke & Burton, 2020; Hastings & Burton, 2008; Ward, Hanner, & Hebert, 2009).

However, recent metabarcoding studies of marine fishes have focused instead on a short segment of the mitochondrial *12S* RNA gene because it provides species-level resolution for many fishes while being vertebrate-specific (Miya et al., 2015; Valsecchi et al., 2019). The smaller *12S* locus is also thought to be advantageous for eDNA studies because of DNA isolated from the environment tends to be degraded and commonly used sequencing technologies target relatively small loci (Collins et al., 2019; Jo et al., 2017; Miya et al., 2015). Given the success of this metabarcoding primer set, the MiFish Universal Teleost primer set is the most commonly used *12S* barcode region because of its utility across a diverse assemblage of marine fishes (Bista et al., 2017; Closek et al., 2019; Thomsen et al., 2016; Valsecchi et al., 2019; Yamamoto et al., 2017).

Thus while there is a near complete *COI* barcode database of California Current fishes (Hastings & Burton, 2008), there is a relative lack of *12S* barcodes for California Current fishes in existing reference databases; GenBank has MiFish *12S* barcodes for 459 of the 864 California fish species (NCBI download October 2019). This paucity of barcodes severely limits the utility of *12S* metabarcoding approaches in California Current coastal waters (Andruszkiewicz et al., 2017; Djurhuus et al., 2020; Port et al., 2015), where relatively recently established marine protected areas (Gleason et al., 2013; Pondella et al., 2015; Thompson, Watson, McClatchie, & Weber, 2012) have created an urgent need for effective and economical monitoring (Elena Maria Duke, Harada, & Burton, 2018; Harada et al., 2015).

Metabarcoding has the ability to help marine resource managers address critical questions, ranging from shifting species distributions, effectiveness of marine protected areas, and seasonal patterns of larval fish recruitment, among others (Closek et al., 2019; Djurhuus et al., 2020; Elena M Duke & Burton, 2020; Kelly, Port, Yamahara, Martone, et al., 2014). However, the success of metabarcoding efforts to enhance fishery management in the California Current Large Marine Ecosystem depends on the development of an improved *12S* barcode reference database. Towards this end, we developed the FishCARD reference database. This regionally-specific database is curated for marine fishes found in the California Current Large Marine Ecosystem, comprised of *12S* sequences previously available in GenBank supplemented by hundreds of additional *12S* sequences generated during this study.

## Methods

### *Reference Barcode Generation*

To generate a more complete *12S* barcode reference database for fish found in the California Current Large Marine Ecosystem, we assembled a list of native marine teleost and elasmobranchs, comprising a total of 864 species (Allen & Horn, 2006; Froese & Pauly, 2010; Hastings & Burton, 2008; Love, & Passarelli, 2020) (Supplemental Table 1). From this list, we acquired as many ethanol-preserved specimens as possible from the Scripps Institution of Oceanography Marine Vertebrates Collection at University of California San Diego (SIO). We obtained a total of 757 samples, representing 612 species (Supplemental Table 2) or 70.8% of all described species of California Current marine fishes. Of these 757 samples, 258 had no previous *12S*

barcodes.

For each sample, we extracted DNA from ~0.25 mg of tissue in 300  $\mu$ L of a 10% Chelex slurry (Walsh, Metzger, & Higuchi, 1991). Given the high volume of samples to process, we initially froze sample slurries at -20°C. Subsequently, samples were thawed, vortexed for 10 seconds, and then centrifuged at high speed for 15 seconds prior to incubating at 95°C for 20 minutes. Samples were then vortexed and centrifuged again at high speed and stored at 4°C until use.

We amplified all DNA extracts using the MiFish Universal Teleost Primers and additionally amplified all elasmobranch samples using the MiFish Elasmobranch Primers (Miya et al., 2015). PCR amplification was conducted following the thermocycler profile of Curd *et al.* (2019b). PCR reactions had 25  $\mu$ L reaction volume containing 12.5  $\mu$ L QIAGEN Multiplex Taq PCR 2x Master Mix (Qiagen Inc., Valencia, CA, USA), 6.5  $\mu$ L of molecular grade water, 2.5  $\mu$ L of each primer (2  $\mu$ mol/L), and 1  $\mu$ L DNA extraction. PCR thermocycling employed a touchdown profile with an initial denaturation at 95°C for 15 min to activate the DNA polymerase, followed by 13 cycles of a 30s denaturation at 94°C, a 30s annealing that started at 69.5°C and then decreased by 1.5°C for each subsequent cycle (last cycle was 50°C), finishing with a 1 min extension at 72°C. This initial touchdown profile was followed by 35 additional cycles using identical parameters except a constant annealing temperature of 50°C and ending with a final extension at 72°C for 10 min. All PCRs included a negative control, where molecular grade water replaced the DNA extraction. All PCR products were visualized via electrophoresis on 2% agarose gels to ensure amplification success and correct product size.

PCR products were purified using ExoSAP-IT (Affymetrix, Cleveland, OH, USA) and sequenced in both directions using BigDye chemistry (Applied Biosystems Inc, Foster City, CA, USA) at Laragen Inc., (Culver City, CA, USA). We trimmed and aligned forward and reverse sequences in Sequencher version 5.4.6 (Nishimura, 2000). All taxonomic names between GenBank and vouchered specimens were synonymized to NCBI taxonomy using the *R* package *taxize* (Chamberlain & Szöcs, 2013). The resulting *12S* sequences were deposited into GenBank (XXXXXX-XXXXXX; Supplemental Table 2).

#### *eDNA Metabarcoding*

To test the utility of the FishCARD database, we metabarcoded fish eDNA from 3 sites off Eastern Santa Cruz Island, CA in 2017. We collected seawater samples from 10m depth using a 4 L Niskin bottle. From this sample, we gravity filtered 1 L through a 0.2  $\mu$ m Sterivex (Millipore Sigma, Burlington, MA, USA) filter in the field and preserved on ice before being transported to a -20°C freezer. We then extracted eDNA from the filters using the modified Qiagen DNeasy Blood and Tissue Kit extraction protocol of Spens et al. (2017), and then amplified it via PCR using the MiFish Teleost primers and thermocycler profile above. We prepared libraries following the methods of Curd et al. (2019) and sequenced these samples on an Illumina MiSeq PE 2x300 at UCLA Technology Center for Genomics and Bioinformatics (see Supplemental Methods for detailed library preparation protocol).

#### *Bioinformatics and Taxonomic Assignment*

We processed DNA sequences using the *Anacapa Toolkit* (Curd et al., 2019) following default parameters with a Bayesian cutoff score of 60. We then assigned taxonomy to each generated amplicon sequence variant (ASV), unique sequence generated through metabarcoding, using three different reference databases. First, we created a *12S* reference database using *CRUX* (Curd et al., 2019), which compiled all publicly available matching *12S* barcode sequences from the NCBI GenBank database targeted by the MiFish Universal Teleost primers, employing standard *CRUX* parameters (Benson et al., 2018; Curd et al., 2019). This set of sequences is herein referred to as the “*CRUX -12S* database” and included any GenBank reference barcode that *in silico* amplified to the MiFish *12S* primers (sequences downloaded in October 2019; <https://github.com/zjgold/FishCARD> & [datadryad.org](http://datadryad.org) link provided upon acceptance). Second, to evaluate how increasing database coverage improves taxonomic assignments, we supplemented the *CRUX-12S* database with the 757 additional California Current fish *12S* barcodes generated for this study, herein referred to as the “combined database”. Third, to test the value of a database curated for the region, we cre-

ated a reference database comprised of only *12S* barcodes of fishes native to California Current Large Marine Ecosystem. These sequences included those obtained from GenBank via *CRUX* and the 757 newly generated reference sequences. This regionally specific reference database is subsequently referred to as “FishCARD”.

### Reference Database Comparisons

To compare the effectiveness of the three reference databases, we examined the total number of ASVs and taxonomic ranks identified using each database. We also investigated differences in taxonomic assignment between single direction ASVs (comprised of forward- and reverse-only sequence reads) and merged ASVs (merged paired-end sequence reads). Specifically, we compared the *CRUX-12S* database to the combined database to determine how the inclusion of additional region-specific reference barcodes improved taxonomic assignment of California Current fish species from eDNA samples. Next, we compared taxonomic assignments between the FishCARD California Current-specific *12S* reference database and the combined reference database to determine whether a curated metabarcoding reference database specific to a regional fauna performs better than a database that uses all available *12S* reference barcodes.

## Results

### Reference Barcode Generation

We generated 757 *12S* MiFish barcodes for 612 California Current fishes, 557 teleosts, 51 elasmobranchs, and 4 cyclostomatan (Supplemental Table 2). Of these, 56 barcodes were duplicates of the same elasmobranch taxa amplified with both the MiFish Elasmobranch and MiFish Universal *12S* primer sets. In total, we generated an additional 258 novel *12S* barcodes for California Current fishes compared to what was available in GenBank at the time of publication. Combining these barcodes with existing reference sequences deposited on GenBank (October 2019) provides reference *12S* barcodes for 715 out of 864 (82.8%) California Current fish species.

### eDNA Metabarcoding

We generated a combined 330,877 sequence reads from 3 eDNA samples, resulting in a total of 2,152 ASVs including singletons; this total dropped to 341 ASVs excluding singletons. Of these 341 ASVs, 211 ASVs were merged paired-reads, 123 ASVs were forward-only reads, and 7 ASVs were reverse only reads. All comparisons below excluded singletons (ASVs that occurred only once across the three samples).

### Unassigned MiFish *12S* ASVs

The *Anacapa Toolkit* failed to assign taxonomy to 49.6% (169/341) of ASVs representing 24.5% (81,002/330,877) of all reads using all three reference databases investigated in this study (Supplemental Table 4). Of the 169 unassigned ASVs, 16 were forward-only reads, and 153 were merged reads. To explore the origins of these unassigned reads, we used BLAST to query all Genbank sequences, revealing that 94.7% (160/169) of these ASVs aligned to marine prokaryotic and eukaryotic *16S* sequences (Max Alignment Scores 87.9-475). Of these aligned ASVs, 85% (136/160) matched to uncultured sequences generated from marine metagenomic studies. 80.0% (128/160) of successfully aligned ASVs matched to bacterial barcodes including those from *Psychromonassp.*, *Photococcus caeruleum*, *Loktanella sp.*, *Leucothrix sp.*, and *Gimesia sp.*, and cyanobacteria. A smaller fraction of assigned ASVs (18.8%; 30/160) best aligned to eukaryotic sequences including those from diatoms (e.g. *Nitzschia alba* and *Eucampia antarctica*) and other marine microalgae (e.g. Picobiliphytes, *Heterosigma akashiwo*, *Mesopedinella arctica*, and *Phacus warszewiczii*). Given that these 169 unassigned sequences were non-vertebrate, we excluded these ASVs from all subsequent comparisons. All remaining 172 ASVs were assigned to a Class of vertebrates by at least one of the three reference databases used. Of these vertebrate ASVs, 58 were merged, 107 were forward-only, and 7 were reverse only reads.

### Reference Database Comparisons

Samples processed using the *CRUX-12S* reference database assigned 89.5% (154/172) of vertebrate ASVs to Family-level and 84.3% (145/172) of vertebrate ASVs to species-level (Supplemental Table 5). Examining

merged reads only, the *CRUX-12S* reference database assigned 94.8% (55/58) of all vertebrate ASVs to Family-level and 87.9% (51/58) of all vertebrate ASVs to species-level (Supplemental Table 6). Examining only forward-reads, the *CRUX-12S* database assigned 87.9% (94/107) of all vertebrate ASVs to Family-level and 83.2% (89/107) of all vertebrate ASVs to species-level. Across all vertebrate ASVs, the *CRUX-12S* database produced 31 unique Family-level assignments, 39 unique genus-level assignments, and 38 unique species-level assignments of which only 25 were species native to the California Current Large Marine Ecosystem. Across all vertebrate ASVs, the *CRUX-12S* database failed to resolve 1.7% (3/172) of vertebrate ASVs to a vertebrate Class, 7% (12/172) of vertebrate ASVs below Class-level, and 1.7% (3/172) of vertebrate ASVs below Order-level.

We then conducted taxonomic assignments using the combined database comprised of both *CRUX-12S* reference barcodes supplemented with the additional 757 barcodes generated in this study. Samples processed using the combined reference database assigned 100% (172/172) of vertebrate ASVs to Family-level and 88.4% (152/172) of vertebrate ASVs to species-level. For merged reads only, the combined database assigned 100% (58/58) of all vertebrate ASVs to Family-level and 93.1% (54/58) of all vertebrate ASVs to species-level. For forward-only reads, the combined database assigned 100% (107/107) of all vertebrate ASVs to Family-level and 87.9% (94/107) of all vertebrate ASVs to species-level. Across all vertebrate ASVs, the combined database produced 28 unique Family-level assignments, 38 unique genus-level assignments, and 37 unique species-level assignments of which 36 were species native to the California Current Large Marine Ecosystem. No ASVs were assigned to only the Class- or Order-level.

Lastly, we assigned taxonomy using the curated FishCARD reference database comprised of only California Current fish *12S* reference barcodes. Samples processed using the curated FishCARD reference database resulted in 99.4% (171/172) of all ASVs assigned to Family-level and 90.7% (156/172) of all ASVs assigned to species-level. For merged reads only, the FishCARD database assigned 100% (58/58) of all vertebrate ASVs to Family-level and 95% (55/58) of all vertebrate ASVs to species-level. For forward-only reads, the FishCARD database assigned 100% (107/107) of all vertebrate ASVs to Family-level and 86.0% (92/107) of all vertebrate ASVs to species-level.

Across all vertebrate ASVs, the FishCARD database produced 27 Family-level assignments, 39 genus-level assignments, and 37 species-level assignments of which all were native CA species. The FishCARD database failed to assign one vertebrate ASV which was assigned to the Family Delphinidae by both the *CRUX-12S* and combined databases.

### *Reference Database Comparisons*

#### *CRUX-12S Database vs. Combined Database*

Comparing the *CRUX-12S* and combined database results demonstrated that the inclusion of novel voucher sequences allowed for species-level identification for 11 additional California Current taxa including Kelp bass (*Paralabrax clathratus*), California moray (*Gymnothorax mordax*), Opaleye (*Girella nigricans*), Giant kelpfish (*Heterostichus rostratus*), Ocean whitefish (*Caulolatilus princeps*), and California halibut (*Paralichthys californicus*) (Supplemental Table 3). The use of the FishCARD reference database also resulted in improved taxonomic classification for 4 species previously missing barcodes that were assigned to other native California Current fish species using the combined database (Supplemental Table 3).

Surprisingly, the inclusion of additional California Current-specific fish *12S* barcodes also led to 10 ASVs receiving a less specific taxonomic classification compared to the *CRUX-12S* database alone. Ten ASVs assigned to the California native Señorita (*Oxyjulis californica*) by the *CRUX-12S* database were only identified to the Family Labridae using FishCARD.

#### *Combined Database vs. Curated Database*

The curated California-specific FishCARD database and the combined *CRUX-12S* reference database supplemented with the additional California-specific *12S* barcodes differed for three taxonomic assignments. First, the curated FishCARD database assigned an ASV to the Black croaker (*Cheilotrema saturnum*)

that was previously only assigned to the family Sciaenidae by the combined database. Second, the curated FishCARD database identified one ASV as native Bat ray (*Myliobatis californica*) whereas the combined database assigned this ASV to the non-native common Eagle ray (*Myliobatis aquila*). Third, FishCARD reference database failed to resolve one ASV previously assigned to the family of Delphinidae by the combined database.

## Discussion

Whether used alone or in combination with existing reference databases, FishCARD reference barcodes dramatically improve the accuracy of eDNA metabarcoding assignments from California Current coastal waters, including species for recreational and commercial fishing and marine ecosystem assessments (Allen & Horn, 2006; Pondella et al., 2015; Sprague et al., 2013). In a test eDNA dataset from 3 sites on Santa Cruz Island, FishCARD performed better, identifying ASVs to species for an additional 15 California Current fishes that were not identified by the *CRUX-12S* database. This increase in accuracy greatly improves the utility of eDNA for monitoring California Current coastal ecosystems, echoing previous research on the importance of complete reference databases in metabarcoding (Leray, Boehm, Mills, & Meyer, 2012; Machida et al., 2017).

Unexpectedly, almost half of the ASVs and a quarter of all sequences generated in our eDNA test datasets could not be assigned to species. While other metabarcoding studies report similar results (Leray & Knowlton, 2017) (Supplemental Table 1), further investigation showed that the vast majority of unassigned ASVs were not fish. Instead, they mapped to uncultured bacteria *16S* loci derived from marine shotgun sequencing metagenomic studies (Bork et al., 2015). This unexpected result highlights that the MiFish Teleost *12S* primer set, while extremely useful for targeting vertebrate *12S* loci, can also amplify non-target *16S* genes, potentially inflating the number of ASVs unassigned to species using this primer set.

### Importance of Complete Reference Databases

Previous eDNA metabarcoding efforts in the California Current reported poor species-level identification and frequent taxonomic assignment to non-native sister taxa (Closek et al., 2019; Kelly, Port, Yamahara, & Crowder, 2014; Port et al., 2015), results that are likely due to the lack of adequate reference sequences for these species. For example, an eDNA metabarcoding study in Southern California (Curd et al., 2019) assigned multiple *12S* ASVs to *Girella simplicidens*, the Gulf opaleye, a fish native to the Gulf of California that does not occur in California coastal waters (Froese & Pauly, 2010; Love & Passarelli, 2020). This incorrect assignment occurred because there were no corresponding *12S* reference sequences for the local native Opaleye, *G. nigricans*. By maximizing the number of reference barcodes from local species, FishCARD allows the vast majority of reads to be correctly assigned to ecologically and geographically relevant species.

Compared to the *CRUX* database generated from *12S* fish sequences in *CRUX-12S*, FishCARD improved species-level assignments, identifying an additional 21.8% of the vertebrate reads in our eDNA samples. Much of this improvement was due to the inclusion of reference barcodes for Kelp bass (*Paralabrax clathratus*), one of the most abundant marine species in Southern California kelp forest ecosystems and an important sport fishery species (Pondella et al., 2015). By including this species, FishCARD assigned 2 previously unidentified ASVs to *P. clathratus*, which accounted for 16.4% of our total sequence reads.

### Importance of regional reference databases

Given that increasing reference database completeness increased the ability to assign ASV's to species, it is logical to assume that databases with more taxonomic coverage are better. However, our results suggest an unexpected trade-off between greater diversity of barcodes and regionally/ecologically informed taxonomic assignment. For example, using only the FishCARD database, which is specific to California Current marine fishes, we identified important native taxa like Black croaker (*Cheilotrema saturnum*) and Bat ray (*Myliobatis californica*) in eDNA samples. However, when FishCard and the *CRUX-12S* databases were combined to yield a database with the largest total number of barcodes, black croaker was not identified and bat ray inconsistently identified across multiple ASVs. The combined database failed to identify black croaker due to the high similarity of *12S* barcode sequences within the Family Sciaenidae, specifically within the clade that

includes *Cheilotrema*, a genus native to California, as well as *Equetus* and *Pareques*, non-native coral reef associated genera; Supplemental Table 3). Similarity of barcode sequences also explains the loss of taxonomic resolution in *Myliobatis*.

By excluding highly similar non-native barcodes, the curated FishCARD database provided more accurate species-level assignments, suggesting that a database comprised of only local taxa is preferred to maximize identification of local species. However, this improvement was not universal. For example, FishCARD failed to classify an ASV belonging to the family Delphinidae that was identified by both the CRUX and combined databases. This result stems from FishCARD being specific to California Current fishes and does not include marine mammals. This shortcoming could be easily overcome, however, by appending FishCARD with barcodes for other marine-associated vertebrate taxa of local management interests (Valsecchi et al., 2019).

These results highlight the tradeoff between identifying local species from clades with little genetic variation and providing taxonomic coverage across a broad range of vertebrate species. As such, researchers need to identify their research priorities when deciding on which reference databases to use, with a particular focus on defining the scope of the target taxa. Future work could alleviate this tradeoff by building bioinformatic pipelines that prioritize assignments to a reference set of native species, perhaps by including information on species ranges and sample locations in the assignment algorithm. Alternatively, a regional database could be appended to address specific questions, such as testing for the presence of specific invasive species or range shifts associated with climate change.

#### *Importance of full-length amplicons*

Although FishCARD typically increased taxonomic resolution of ASVs from our eDNA samples, in one specific case the inclusion of additional California Current-specific *12S* barcodes led to substantially reduced taxonomic resolution, with 10 ASVs initially assigned to *Oxyjulis californica* subsequently assigned only to family, Labridae. This result appears to suggest that inclusion of barcodes from local fauna decreases assignment accuracy. However, all 10 of these ASVs were forward sequences only. The Señorita (*Oxyjulis californica*) and Rock wrasse (*Halichoeres semicinctus*), both native to the California Current share a high degree of similarity in the forward *12S* sequence fragment, and the diagnostic SNPs occur in the reverse sequence fragment. This issue is resolved when using the full-length merged reads which capture diagnostic SNPs between these two species. Given genetic similarities between these two species, a recent phylogenetic study of new world *Halichoeres* wrasses has even argued for synonymizing *Oxyjulis* into *Halichoeres* (Wainwright et al., 2018).

Despite the above, single direction reads can provide accurate taxonomic assignment for many California Current fishes. For example, within Labridae, the same family as Señorita and Rock wrasse, forward-reads provided species-level resolution for the California sheephead, *Semicossyphus pulcher*. Such variation in taxonomic assignment based on single-direction sequences highlights the importance of full-length merged reads. To ensure the most accurate taxonomic classifications, researchers should maximize the generation of merged reads by choosing the appropriate sequencing platform (e.g. paired end 300 bp vs. paired end 75 bp).

Although we conducted 300 bp paired-end high throughput sequencing for a 176 bp amplicon, we still obtained a substantial number of orphaned single end reads. This result is largely due to the generation of low quality reverse sequence reads, a persistent issue found across Illumina sequencing platforms (Callahan et al., 2016; Kwon, Park, Lee, & Yoon, 2013). As such, it is important to think carefully about whether or not to include single end reads on a case-by-case and potentially species-by-species basis. In the present study, we were only able to make this determination because of the relative completeness of the FishCARD database, which included all labrids native to the California Current.

#### *Limitations of FishCARD*

The FishCARD database did not include barcodes for all California Current marine fishes due to a combination of limited resources, difficulties amplifying vouchered tissue samples, and a lack of some vouchered reference material within the Marine Vertebrates Collection of the Scripps Institution of Oceanography. Of

the 149 (17.2%) California Current fishes absent from FishCARD, 44.0% (n=66) are rare in the California Current, 14.7% (n=22) are common but not coastal species, and 2.0% (n=3) were introduced estuarine species; only, 18.0% (n=27) were common coastal species (Supplemental Table 7). As such, FishCARD provides coverage for the vast majority of California Current marine fishes, making it an important tool for metabarcoding studies, despite these missing taxa.

The one major shortcoming of FishCARD is that 20.8% (n=31) of the missing taxa are in the Genus *Sebastes*, and rockfish are ecologically important (Hyde & Vetter, 2007), form the basis of many commercial and recreational fisheries (Lea, McAllister, & VenTresca, 1999; Williams, Levin, & Palsson, 2010), and declines in rockfish stocks led to the establishment of the largest marine protected areas in southern California, the Cowcod Conservation Areas (Thompson et al., 2017). Unfortunately, this shortcoming cannot be easily overcome through additional *12S* barcoding. This is because rockfish are a recent and diverse radiation comprised of 110 species (Ingram & Kai, 2014) and *12S* fails to resolve most *Sebastes* to species-level (Hyde & Vetter, 2007; Yamamoto et al., 2017). Thus effective metabarcoding of *Sebastes* will require designing novel *Sebastes*- specific metabarcoding primers that target a more rapidly evolving region of the mitochondrial genome (e.g. *CytB*) (Thompson et al., 2017).

However, FishCARD includes 100% of all non-*Sebastes* nearshore species monitored by the Channel Islands National Kelp Forest Monitoring Program (n=80, Sprague et al., 2013), as well as by PISCO, the Partnership for Interdisciplinary Studies of Coastal Oceans (n=76; Caselle, Rassweiler, Hamilton, & Warner, 2015; Pondella et al., 2015). Further, there is now *12S* reference sequence for 98 of the 100 most abundant ichthyoplankton species collected by the California Cooperative Oceanic Fisheries Investigation (CalCOFI) from the California Current between 1951-2019 (only Showy bristlemouth *Cyclothone signata* and Spotted barracudina *Arctozenus risso*) (Moser, 1993). Moreover, in real world application, this reference barcode database assigned taxonomy to over 90% of vertebrate ASVs detecting a broad range of ecologically and commercially important nearshore rocky reef species (Pondella II et al., 2019). As such, FishCARD represents an important genetic resource for coastal California marine metabarcoding monitoring efforts.

#### *Limitations of MiFish primers*

High numbers of unidentified ASVs are a common feature of barcoding and metabarcoding studies (e.g. Leray & Knowlton, 2017). These unidentified ASVs are typically attributed to incomplete reference databases (Curd et al., 2019; Ransome et al., 2017; Schenekar et al., 2020) and/or novel biodiversity (Barber & Boyce, 2006; Boussarie et al., 2018). However, given that FishCARD includes 82.8% of all California Current fishes, and the remaining 15.6% are mostly rare species unlikely to be found in a kelp forest, it was extremely surprising that half of all ASVs and a quarter of all sequences generated in our eDNA test datasets could not be assigned.

The vast majority of these sequences and ASVs did not belong to vertebrates, but instead uncultured marine bacteria, specifically matching to *16S*, rather than *12S* loci. Since mitochondria represent the capture of microbial endosymbionts by ancient eukaryotes (Roger, Muñoz-Gómez, & Kamikawa, 2017) and that this capture occurred in the sea, it perhaps is not surprising that primers designed to target vertebrate *12S* might also capture marine prokaryotes. Similarly, the homology between vertebrate *12S* and prokaryotic and bacterial *16S* genes is well known (Crews & Attardi, 1980) suggesting capturing microbial *16S* with vertebrate *12S* primers is also not surprising. However, this particular feature of the MiFish primer set was previously unreported, potentially impacting the interpretation of unidentified ASVs in fish metabarcoding studies.

These findings highlight the importance of accurate universal metabarcoding primer design, especially in outlining both target and non-target sequences. In the design of the MiFish Teleost *12S* primers, uncultured marine microbe *16S* sequences were not considered as potential alternative targets for the primer set, resulting in the selection of a metabarcoding locus with a high degree of non-target amplification (Miya et al., 2015). This is an important finding for the marine vertebrate eDNA community which has recently converged on the MiFish *12S* primers (Closek et al., 2019; O'Donnell et al., 2017; Valsecchi et al., 2019; Yamahara

et al., 2019) as these results suggest that the MiFish *12S* primer set will generate substantial quantities of non-target eDNA reads. At best, this non-target amplification will lead wasted sequencing effort, as every microbial sequence generated reduces the number of vertebrate sequences. At worst, it could result in incorrect interpretation of unidentified ASVs. This problem is of particular concern in environments with high relative abundance of marine bacterial communities and low relative abundance of vertebrate biomass such as in some pelagic midwater and deep sea habitats where recent eDNA sample collection efforts have struggled to detect vertebrate sequences (K. Pitz personal communication).

### *Towards improved metabarcoding*

FishCARD was designed to improve effectiveness of metabarcoding of California Current marine fishes. To further improve and expand the taxonomic coverage of the database, we generated a website that identifies species needing *12S* reference barcodes and provides the research community targets for additional barcoding efforts (<https://github.com/zjgold/FishCARD>). The ability to update and expand FishCARD will be especially important as climate change leads to range expansions of sub-tropical species that may become resident within the California Current Large Marine Ecosystem (Gentemann, Fewings, & García-Reyes, 2017; Harvell et al., 2019; Sanford, Sones, Garcia-Reyes, Goddard, & Largier, 2019). The importance of expanding the database is highlighted by our detection of Finescale triggerfish, *Balistes polylepis*, in the eDNA samples, a species that has only recently become more common off Santa Cruz Island and La Jolla since the 2014-2016 marine heatwave (B. Frable & S. McMillan, personal communication).

Additionally while the MiFish Teleost and Elasmobranch *12S* loci are important targets for current marine metabarcoding studies, future efforts and different applications of marine metabarcoding will likely rely on additional barcoding targets. Recent efforts have found success multiplexing *CO1* and *16S* loci simultaneously, providing more species-level identifications than either marker alone and demonstrating complimentary genetic loci can improve metabarcoding assignments (Duke & Burton, 2020). Future efforts to develop rapid and affordable multi-loci barcoding and mitogenomic tools will provide greater resources for marine metabarcoding and population genomic efforts (Coissac, Hollingsworth, Lavergne, & Taberlet, 2016). As these new barcode loci are developed (e.g. *Sebastes* -specific barcodes), FishCARD can be expanded to include these loci. Additionally, resources like the SIO Marine Vertebrates Collection will continue to provide important voucher specimens for advancing marine molecular ecology resources as they accession new material.

Here we demonstrate that FishCARD provides an important genetic resource for California Current marine metabarcoding efforts, improving the accuracy and effectiveness of this important and growing research tool. The development of robust and complete reference databases dramatically improves the accuracy of species-level taxonomic assignments, in turn enhancing the efficacy and applicability of these tools for marine biomonitoring. This tool dramatically improves fish eDNA metabarcoding efforts in the California Current Large Marine Ecosystem and provides marine resource managers and researchers an important tool for surveying and monitoring marine fish communities using eDNA.

### **Acknowledgements**

We would like to acknowledge NSF GRFP and GRIP [DEG No. 2015204395], UCLA La Kretz Center for Conservation Genomics, and UCLA Department of Ecology and Evolutionary Biology for funding this research. In addition, we would like to thank Regina Wetzler, Dean Pentcheff, Adam Wall, Janessa Wall, and Giacomo Bernardi for providing additional voucher specimens. We would like to thank Julianne K. Passarelli and Milton S. Love for sharing their comprehensive list of California fish species from their new edition of *Miller and Lea's Guide to the Coastal Marine Fishes of California* (Love & Passarelli, 2020). Lastly, we would like to thank McKenzie Koch, Beverly Shih, Nikita Sridhar, Lauren Man, Eric Caldera, and Erick Zerecero for help conduct all laboratory work. We would like to thank William Watson, Phillip A. Hastings, Peggy Fong, Richard F. Ambrose, and Thomas B. Smith for providing thoughtful feedback and comments during the preparation of the manuscript.

### **References**

- Allen, L. G., & Horn, M. H. (2006). *The ecology of marine fishes: California and adjacent waters*. Univ of California Press.
- 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* (4), e0176343. doi:10.1371/journal.pone.0176343
- Ardura, A., Planes, S., & Garcia-Vazquez, E. (2013). Applications of DNA barcoding to fish landings: authentication and diversity assessment. *ZooKeys*, (365), 49–65. doi:10.3897/zookeys.365.6409
- Baetscher, D. S., Anderson, E. C., Gilbert-Horvath, E. A., Malone, D. P., Saarman, E. T., Carr, M. H., & Garza, J. C. (2019). Dispersal of a nearshore marine fish connects marine reserves and adjacent fished areas along an open coast. *Molecular Ecology*, *28* (7), 1611–1623.
- Barber, P., & Boyce, S. L. (2006). Estimating diversity of Indo-Pacific coral reef stomatopods through DNA barcoding of stomatopod larvae. *Proceedings of the Royal Society B: Biological Sciences*, *273* (1597), 2053–2061.
- Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Ostell, J., Pruitt, K. D., & Sayers, E. W. (2018). GenBank. *Nucleic Acids Research*, *46* (D1), D41–D47.
- Bista, I., Carvalho, G. R., Walsh, K., Seymour, M., Hajibabaei, M., Lallias, D., ... Creer, S. (2017). Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. *Nature Communications*, *8*, 14087.
- Bork, P., Bowler, C., De Vargas, C., Gorsky, G., Karsenti, E., & Wincker, P. (2015). Tara Oceans studies plankton at planetary scale. American Association for the Advancement of Science.
- Boussarie, G., Bakker, J., Wangensteen, O. S., Mariani, S., Bonnin, L., Juhel, J.-B., ... Robbins, W. D. (2018). Environmental DNA illuminates the dark diversity of sharks. *Science Advances*, *4* (5), eaap9661.
- Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). obitools: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, *16* (1), 176–182.
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13* (7), 581.
- Caselle, J. E., Rassweiler, A., Hamilton, S. L., & Warner, R. R. (2015). Recovery trajectories of kelp forest animals are rapid yet spatially variable across a network of temperate marine protected areas. *Scientific Reports*, *5*, 14102. doi:10.1038/srep14102
- Chamberlain, S. A., & Szocs, E. (2013). taxize: taxonomic search and retrieval in R. *F1000Research*, *2*.
- Closek, C. J., Santora, J. A., Starks, H. A., Schroeder, I. D., Andruszkiewicz, E. A., Sakuma, K. M., ... Boehm, A. B. (2019). Marine vertebrate biodiversity and distribution within the central California Current using environmental DNA (eDNA) metabarcoding and ecosystem surveys. *Frontiers in Marine Science*, *6*, 732.
- Coissac, E., Hollingsworth, P. M., Lavergne, S., & Taberlet, P. (2016). From barcodes to genomes: extending the concept of DNA barcoding. *Molecular Ecology*, *25* (7), 1423–1428.
- Collins, R. A., Bakker, J., Wangensteen, O. S., Soto, A. Z., Corrigan, L., Sims, D. W., ... Mariani, S. (2019). Non-specific amplification compromises environmental DNA metabarcoding with COI. *Methods in Ecology and Evolution*, *10* (11), 1985–2001. doi:10.1111/2041-210X.13276
- Crews, S., & Attardi, G. (1980). The sequences of the small ribosomal RNA gene and the phenylalanine tRNA gene are joined end to end in human mitochondrial DNA. *Cell*, *19* (3), 775–784.

- Curd, E. E., Gold, Z., Kandlikar, G. S., Gomer, J., Ogden, M., O'Connell, T., ... Meyer, R. S. (2019). Anacapa: an environmental DNA toolkit for processing multilocus metabarcode datasets. *Methods in Ecology and Evolution* , 10 , 1469–1475. doi:<https://doi.org/10.1111/2041-210X.13214>
- Deiner, K., Bik, H. M., Machler, E., Seymour, M., Lacoursiere-Roussel, A., Altermatt, F., ... Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology* , 26 (21), 5872–5895. doi:10.1111/mec.14350
- Djurhuus, A., Closek, C. J., Kelly, R. P., Pitz, K. J., Michisaki, R. P., Starks, H. A., ... Hubbard, K. (2020). Environmental DNA reveals seasonal shifts and potential interactions in a marine community. *Nature Communications* , 11 (1), 1–9.
- Duke, Elena M, & Burton, R. S. (2020). Efficacy of metabarcoding for identification of fish eggs evaluated with mock communities. *Ecology and Evolution* , 10 (7), 3463–3476. doi:10.1002/ece3.6144
- Duke, Elena Maria, Harada, A. E., & Burton, R. S. (2018). Large interannual variation in spawning in San Diego marine protected areas captured by molecular identification of fish eggs. *Marine Ecology Progress Series* , 604 , 199–210.
- Froese, R., & Pauly, D. (2010). FishBase. Fisheries Centre, University of British Columbia.
- Gentemann, C. L., Fewings, M. R., & Garcia-Reyes, M. (2017). Satellite sea surface temperatures along the West Coast of the United States during the 2014–2016 northeast Pacific marine heat wave. *Geophysical Research Letters* , 44 (1), 312–319.
- Gleason, M., Fox, E., Ashcraft, S., Vasques, J., Whiteman, E., Serpa, P., ... Wiseman, K. (2013). Designing a network of marine protected areas in California: Achievements, costs, lessons learned, and challenges ahead. *Ocean & Coastal Management* , 74 , 90–101. doi:10.1016/j.ocecoaman.2012.08.013
- Guo, J. (2017). Metabarcoding Analyses of Gut Microbiome Compositions in Red Abalone (*Haliotis Rufescens*, Swainson, 1822) Fed Different Macroalgal Diets.
- Harada, A. E., Lindgren, E. A., Hermsmeier, M. C., Rogowski, P. A., Terrill, E., & Burton, R. S. (2015). Monitoring Spawning Activity in a Southern California Marine Protected Area Using Molecular Identification of Fish Eggs. *PLOS ONE* , 10 (8), e0134647. Retrieved from <https://doi.org/10.1371/journal.pone.0134647>
- Harvell, C. D., Montecino-Latorre, D., Caldwell, J. M., Burt, J. M., Bosley, K., Keller, A., ... Pontier, O. (2019). Disease epidemic and a marine heat wave are associated with the continental-scale collapse of a pivotal predator (*Pycnopodia helianthoides*). *Science Advances* , 5 (1), eaau7042.
- Hastings, P. A., & Burton, R. S. (2008). Establishing a DNA sequence database for the marine fish fauna of California.
- Hyde, J. R., & Vetter, R. D. (2007). The origin, evolution, and diversification of rockfishes of the genus *Sebastes* (Cuvier). *Molecular Phylogenetics and Evolution* , 44 (2), 790–811. doi:10.1016/j.ympev.2006.12.026
- Ingram, T., & Kai, Y. (2014). The geography of morphological convergence in the radiations of Pacific *Sebastes* rockfishes. *The American Naturalist* , 184 (5), E115–31. doi:10.1086/678053
- Jo, T., Murakami, H., Masuda, R., Sakata, M. K., Yamamoto, S., & Minamoto, T. (2017). Rapid degradation of longer DNA fragments enables the improved estimation of distribution and biomass using environmental DNA. *Molecular Ecology Resources* , 17 (6), e25–e33.
- Kelly, R. P., Port, J. A., Yamahara, K. M., & Crowder, L. B. (2014). Using environmental DNA to census marine fishes in a large mesocosm. *PloS One* , 9 (1), e86175. doi:10.1371/journal.pone.0086175
- Kelly, R. P., Port, J. a., Yamahara, K. M., Martone, R. G., Lowell, N., Thomsen, P. F., ... Crowder, L. B. (2014). Harnessing DNA to improve environmental management. *Science* , 344 (6191). doi:10.1126/science.1251156

- Klymus, K. E., Marshall, N. T., & Stepien, C. A. (2017). Environmental DNA (eDNA) metabarcoding assays to detect invasive invertebrate species in the Great Lakes. *PLOS ONE* , 12 (5), e0177643. Retrieved from <https://doi.org/10.1371/journal.pone.0177643>
- Kwon, S., Park, S., Lee, B., & Yoon, S. (2013). In-depth analysis of interrelation between quality scores and real errors in Illumina reads. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 635–638). IEEE.
- Lea, R. N., McAllister, R. D., & VenTresca, D. A. (1999). *Biological aspects of nearshore rockfishes of the genus Sebastes from central California: with notes on ecologically related sport fishes* (Vol. 177). State of California, The Resources Agency, Department of Fish and Game.
- Leray, M., Boehm, J. T., Mills, S. C., & Meyer, C. P. (2012). Moorea BIOCOTE barcode library as a tool for understanding predator–prey interactions: insights into the diet of common predatory coral reef fishes. *Coral Reefs* , 31 (2), 383–388. doi:10.1007/s00338-011-0845-0
- Leray, M., & Knowlton, N. (2017). Random sampling causes the low reproducibility of rare eukaryotic OTUs in Illumina COI metabarcoding. *PeerJ* , 5 , e3006.
- Love, M. S., & Passarelli, J. K. (2020). *Miller and Lea's Guide to the Coastal Marine Fishes of California* . University of California Agriculture and Natural Resources. Retrieved from <https://books.google.com/books?id=Puc7ywEACAAJ>
- Machida, R. J., Leray, M., Ho, S.-L., & Knowlton, N. (2017). Metazoan mitochondrial gene sequence reference datasets for taxonomic assignment of environmental samples. *Scientific Data* , 4 , 170027.
- Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., ... 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 (7), 150088. doi:10.1098/rsos.150088
- Moser, H. G. (1993). *Distributional atlas of fish larvae and eggs in the California Current region: taxa with 1000 or more total larvae, 1951 through 1984* . Marine Life Research Program, Scripps Institution of Oceanography.
- Nishimura, D. (2000). Sequencher 3.1. 1. *Biotech Software & Internet Report* , 1 (1–2), 24–30.
- O'Donnell, J. L., Kelly, R. P., Shelton, A. O., Samhuri, J. F., Lowell, N. C., & Williams, G. D. (2017). Spatial distribution of environmental DNA in a nearshore marine habitat. *PeerJ* , 5 , e3044. doi:10.7717/peerj.3044
- Pondella, D. J., Caselle, J. E., Claisse, J. T., Williams, J. P., Davis, K., Williams, C. M., & Zahn, L. A. (2015). Baseline Characterization of the Shallow Rocky Reef and Kelp Forest Ecosystems of the South Coast Study Region. Retrieved from <https://caseagrants.ucsd.edu/news/summaries-of-projects-selected-for-funding-through-the-south-coast-mpa-baseline-program>
- Pondella II, D. J., Piacenza, S. E., Claisse, J. T., Williams, C. M., Williams, J. P., Zellmer, A. J., & Caselle, J. E. (2019). Assessing drivers of rocky reef fish biomass density from the Southern California Bight. *Marine Ecology Progress Series* , 628 , 125–140.
- Port, J. A., O'Donnell, J. L., Romero-Maraccini, O. C., Leary, P. R., Litvin, S. Y., Nickols, K. J., ... Kelly, R. P. (2015). Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular Ecology* .
- Ransome, E., Geller, J. B., Timmers, M., Leray, M., Mahardini, A., Sembiring, A., ... Meyer, C. P. (2017). The importance of standardization for biodiversity comparisons: A case study using autonomous reef monitoring structures (ARMS) and metabarcoding to measure cryptic diversity on Mo'orea coral reefs, French Polynesia. *PloS One* , 12 (4).

- Roger, A. J., Munoz-Gomez, S. A., & Kamikawa, R. (2017). The origin and diversification of mitochondria. *Current Biology* , 27 (21), R1177–R1192.
- Sanders, J. G., Beichman, A. C., Roman, J., Scott, J. J., Emerson, D., McCarthy, J. J., & Girguis, P. R. (2015). Baleen whales host a unique gut microbiome with similarities to both carnivores and herbivores. *Nature Communications* , 6 , 8285.
- Sanford, E., Sones, J. L., Garcia-Reyes, M., Goddard, J. H. R., & Largier, J. L. (2019). Widespread shifts in the coastal biota of northern California during the 2014–2016 marine heatwaves. *Scientific Reports* , 9 (1), 4216.
- Schenekar, T., Schletterer, M., Lecaudey, L. A., & Weiss, S. J. (2020). Reference databases, primer choice, and assay sensitivity for environmental metabarcoding: Lessons learnt from a re-evaluation of an eDNA fish assessment in the Volga headwaters. *River Research and Applications* .
- Spens, J., Evans, A. R., Halfmaerten, D., Knudsen, S. W., Sengupta, M. E., Mak, S. S. T., ... Hellstrom, M. (2017). Comparison of capture and storage methods for aqueous microbial eDNA using an optimized extraction protocol: advantage of enclosed filter. *Methods in Ecology and Evolution* , 8 (5), 635–645. doi:10.1111/2041-210X.12683
- Sprague, J. L., Traiger, S. B., Grunden, J. R., Ibarra, S. N., Mooney, E. A., Moore, K. J., & Kushner, D. J. (2013). *Channel Islands National Park kelp forest monitoring program: Annual report 2011* . Fort Collins, Colorado.
- Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. *Molecular Ecology* , 21 (8), 1789–1793. doi:10.1111/j.1365-294X.2012.05542.x
- Thompson, A. R., Chen, D. C., Guo, L. W., Hyde, J. R., & Watson, W. (2017). Larval abundances of rockfishes that were historically targeted by fishing increased over 16 years in association with a large marine protected area. *Royal Society Open Science* , 4 (9), 170639. doi:10.1098/rsos.170639
- Thompson, A. R., Watson, W., McClatchie, S., & Weber, E. D. (2012). Multi-Scale Sampling to Evaluate Assemblage Dynamics in an Oceanic Marine Reserve. *PLoS ONE* , 7 (3), e33131. doi:10.1371/journal.pone.0033131
- Thomsen, P. F., Møller, P. R., Sigsgaard, E. E., Knudsen, S. W., Jørgensen, O. A., & Willerslev, E. (2016). Environmental DNA from seawater samples correlate with trawl catches of subarctic, deepwater fishes. *PLoS ONE* , 11 (11). doi:10.1371/journal.pone.0165252
- Valsecchi, E., Bylemans, J., Goodman, S. J., Lombardi, R., Carr, I., Castellano, L., ... Galli, P. (2019). Novel Universal Primers for Metabarcoding eDNA Surveys of Marine Mammals and Other Marine Vertebrates. *BioRxiv* , 759746.
- Wainwright, P. C., Santini, F., Bellwood, D. R., Robertson, D. R., Rocha, L. A., & Alfaro, M. E. (2018). Phylogenetics and geography of speciation in New World Halichoeres wrasses. *Molecular Phylogenetics and Evolution* , 121 , 35–45. doi:https://doi.org/10.1016/j.ympev.2017.12.028
- Walsh, P. S., Metzger, D. A., & Higuchi, R. (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* , 10 (4), 506–513.
- Ward, R. D., Hanner, R., & Hebert, P. D. N. (2009). The campaign to DNA barcode all fishes, FISH-BOL. *Journal of Fish Biology* , 74 (2), 329–356. doi:10.1111/j.1095-8649.2008.02080.x
- Williams, G. D., Levin, P. S., & Palsson, W. A. (2010). Rockfish in Puget Sound: An ecological history of exploitation. *Marine Policy* , 34 (5), 1010–1020.
- Yamahara, K. M., Preston, C. M., Birch, J. M., Walz, K. R., Marin III, R., Jensen, S., ... Ryan, J. (2019). In-situ Autonomous Acquisition and Preservation of Marine Environmental DNA Using an Autonomous Underwater Vehicle. *Frontiers in Marine Science* , 6 , 373.

Yamamoto, S., Masuda, R., Sato, Y., Sado, T., Araki, H., Kondoh, M., ... Miya, M. (2017). Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea. *Scientific Reports* , 7 , 40368.

### **Data Accessibility**

Reference databases and metabarcoding data is made publicly available and stored on dryad ([datadryad.org](http://datadryad.org) link provided upon acceptance). All reference barcode sequences have been uploaded to GenBank (accession numbers provided upon acceptance). Additional supporting information is available at <https://github.com/zjgold/FishCARD>.

### **Author Contributions**

- \* Conceptualization ZG, ESC, DK, BF, RSB, KDG, ART, PHB
- \* Performed Research ZG, ESC, DK, BF, ART
- \* Funding Acquisition ZG, PHB, ART, KDG, DK, RSB
- \* Data Curation ZG, ESC, DK, BF
- \* Formal Analysis ZG
- \* Writing – Original Draft Preparation ZG, ESC, DK, BF, RSB, KDG, ART, PHB

### **Conflicts of Interest**

The authors have no conflicts of interest to report.

### **Tables**

NA

### **Figures**

NA