Evaluation of DNA metabarcoding for identifying fish eggs: a case study on the West Florida Shelf (#81223)

First submission

Guidance from your Editor

Please submit by **29 Jan 2023** for the benefit of the authors (and your token reward) .

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Custom checks
Make sure you include the custom checks shown below, in your review.

Author notes
Have you read the author notes on the guidance page?

Raw data check
Review the raw data.

Image check
Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

Files

Download and review all files from the materials page.

1 Figure file(s)
3 Table file(s)

Custom checks

DNA data checks

- Have you checked the authors data deposition statement?
- Can you access the deposited data?
- Has the data been deposited correctly?
- Is the deposition information noted in the manuscript?

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- All underlying data have been provided; they are robust, statistically sound, & controlled.
- Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult. I suggest you have a colleague who is proficient in English and familiar with the subject matter review your manuscript, or contact a professional editing service.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

Evaluation of DNA metabarcoding for identifying fish eggs: a case study on the West Florida Shelf

Mya Breitbart [Corresp., 1], **Makenzie Kerr** [1], **Michael J Schram** [1], **Ian Williams** [1], **Grace Koziol** [1], **Ernst Peebles** [1], **Christopher D Stallings** [1]

[1] College of Marine Science, University of South Florida, Saint Petersburg, Florida, United States

Corresponding Author: Mya Breitbart
Email address: mya@usf.edu

A critical factor in fisheries management is the protection of spawning sites for ecologically and economically important fish species. DNA barcoding (i.e., amplification and sequencing of the mitochondrial cytochrome c oxidase I (COI) gene) of fish eggs has emerged as a powerful technique for identifying spawning sites. However, DNA barcoding of individual fish eggs is time-consuming and expensive. In an attempt to reduce costs and effort for long-term fisheries monitoring programs, here we used DNA metabarcoding, in which DNA is extracted and amplified from a composited sample containing all the fish eggs collected at a given site, to identify fish eggs from 49 stations on the West Florida Shelf. A total of 37 taxa were recovered from 4,719 fish eggs. Egg distributions on the West Florida Shelf corresponded with the known habitat types occupied by these taxa, which included burrowers, coastal pelagic, epipelagic, mesopelagic, demersal, deep demersal, commensal, and reef-associated taxa. Metabarcoding of fish eggs was faster and far less expensive than barcoding individual eggs; however, this method cannot provide absolute taxon proportions due to variable copy numbers of mitochondrial DNA in different taxa, different numbers of cells within eggs depending on developmental stage, and PCR amplification biases. In addition, some samples yielded sequences from more taxa than the number of eggs present, demonstrating the presence of contaminating DNA and requiring the application of a threshold proportion of sequences required for counting a taxon as present. Finally, we review the advantages and disadvantages of using metabarcoding versus individual fish egg barcoding for long-term monitoring programs.

Evaluation of DNA metabarcoding for identifying fish eggs: a case study on the West Florida Shelf

Mya Breitbart[1*], Makenzie Kerr[1], Michael J. Schram[1], Ian Williams[1], Grace Koziol[1], Ernst Peebles[1], Christopher D. Stallings[1]

[1]College of Marine Science, University of South Florida, Saint Petersburg, FL, USA

Corresponding Author:
Mya Breitbart
140 7th Ave S, St. Petersburg, Florida, 33701, USA
Email address: mya@usf.edu

Abstract

A critical factor in fisheries management is the protection of spawning sites for ecologically and economically important fish species. DNA barcoding (i.e., amplification and sequencing of the mitochondrial cytochrome c oxidase I (COI) gene) of fish eggs has emerged as a powerful technique for identifying spawning sites. However, DNA barcoding of individual fish eggs is time-consuming and expensive. In an attempt to reduce costs and effort for long-term fisheries monitoring programs, here we used DNA metabarcoding, in which DNA is extracted and amplified from a composited sample containing all the fish eggs collected at a given site, to identify fish eggs from 49 stations on the West Florida Shelf. A total of 37 taxa were recovered from 4,719 fish eggs. Egg distributions on the West Florida Shelf corresponded with the known habitat types occupied by these taxa, which included burrowers, coastal pelagic, epipelagic, mesopelagic, demersal, deep demersal, commensal, and reef-associated taxa. Metabarcoding of fish eggs was faster and far less expensive than barcoding individual eggs; however, this method cannot provide absolute taxon proportions due to variable copy numbers of mitochondrial DNA in different taxa, different numbers of cells within eggs depending on developmental stage, and PCR amplification biases. In addition, some samples yielded sequences from more taxa than the number of eggs present, demonstrating the presence of contaminating DNA and requiring the application of a threshold proportion of sequences required for counting a taxon as present. Finally, we review the advantages and disadvantages of using metabarcoding versus individual fish egg barcoding for long-term monitoring programs.

Introduction

A critical factor in fisheries management is the protection of spawning sites for ecologically and economically important fish species. Studies commonly hindcast spawning sites based on the locations where larvae from a given species have been identified, but this method is imprecise because larvae can be days, weeks, or even months old at the time of capture (Cowen and Sponaugle 2009). In contrast, predicting spawning sites based on the presence of eggs is much more reliable since eggs behave as relatively passive particles (the exception being the eggs of species that are not neutrally buoyant), most fish remain in this developmental stage for a maximum of 1-2 days, and the development stage of the eggs can be determined, allowing identification of eggs that are less than a few hours old, if so desired. However, since fish eggs are difficult to identify visually, the spawning locations of many broadcast spawning species remain unknown (Kawakami et al. 2010, Becker et al. 2015). DNA barcoding (i.e., amplification and sequencing of the mitochondrial cytochrome c oxidase I (COI) gene) of individual fish eggs has emerged as a powerful technique for identification of fish spawning sites (Shao et al. 2002, Saitoh et al. 2009, Lelièvre et al. 2012, Burghart et al. 2014, Frantine-Silva et al. 2015, Harada et al. 2015, Lewis et al. 2016, Leyva-Cruz et al. 2016, Lin et al. 2016, Hofmann et al. 2017, Ahern et al. 2018, Duke et al. 2018, Burrows et al. 2019, Hou et al. 2020, Kerr et al. 2020, Mateos-Rivera et al. 2020, Hou et al. 2022).

Through several pilot studies and the long-term Spawning Habitat & Early-life Linkages to Fisheries (SHELF) program funded by the Florida RESTORE Act Center of Excellence Program (FLRACEP), we have used DNA barcoding to identify over 8,500 individual fish eggs from over 320 locations in the Gulf of Mexico (GOM) and Florida Straits in the past decade (Burghart et al. 2014, Burrows et al. 2019, Keel et al. 2022, Kerr et al. 2020, Kerr et al. 2022). These data have provided tremendous insight into the spatial distribution of fish early life stages in this region, provided the first known spawning grounds for several taxa (Kerr et al. 2020), demonstrated a disparity between the composition of co-occurring egg and larval communities (Burghart et al. 2014), identified distinct distributions of eggs from neritic versus oceanic taxa with a community transition at the shelf break (Burrows et al. 2019), and documented the potential of mesoscale cyclic eddies to entrain the eggs of reef-associated taxa and transport them to deeper waters (Kerr et al. 2020). Building a long-term time series of fish egg community composition is valuable for detecting changes in fish-egg community composition over time; however, DNA barcoding of individual fish eggs is expensive and time-consuming. Two recent studies have demonstrated the use of metabarcoding, in which DNA is extracted and amplified from an aggregate sample containing all the fish eggs collected at a given site, to characterize the spawning community (Duke and Burton 2020, Miranda-Chumacero et al. 2020). Here, we sought to evaluate the performance of metabarcoding as a potential way to increase throughput and reduce both financial and human resource costs for the purposes of supporting a long-term fish egg monitoring program. We applied DNA metabarcoding to identify 4,719 fish eggs collected from 49 samples on the West Florida Shelf, and recovered eggs from a total of 37 taxa.

Materials & Methods

Sample collection

We collected planktonic fish eggs from an a priori-defined grid on the West Florida Shelf (Fig. 1A) aboard the R/V *Hogarth* on two cruises (August 6-16, 2019 and September 24-26, 2019; Supplemental Table 1). At each station, we performed a 15-minute tow at the ocean surface with a bongo (double conical) plankton net (333 µm mesh, 61 cm mouth diameter) equipped with plastic, 1-liter cod-ends, and a General Oceanics 2030R mechanical flowmeter. After each tow, we washed down the nets with ambient seawater. The right-hand cod-ends were drained of excess seawater using a sieve and rinsed back into the jar using 95% isopropanol, leaving the final concentration >50% isopropanol. In the case of high biomass, we split the samples into two cod-ends to allow proper alcohol preservation. We stored the samples at 4°C until processing.

In the laboratory, we picked at least 100 percomorph eggs (except when less were present) per sample using a stereomicroscope, gridded petri dishes, and fine-tipped forceps during the month of December 2019. When fewer than 100 eggs were present, we processed the entire sample. Each sample was separated into a labeled 1.5 mL screw cap, o-ring tube with 70% isopropanol for genetic identification. The number of eggs processed for each sample ranged from 2 – 272 (Supplemental Table 1).

DNA extraction, PCR, sequencing

To start the extraction process, we carefully removed the isopropanol with a sterile pipette tip. Next, we rinsed the eggs twice with molecular grade PCR water to remove any additional alcohol or other contaminants. To break open the chorion of the eggs, we added 0.4 g of 1 mm beads to dry eggs along with 500 µl of HotSHOT alkaline lysis buffer (25 mM NaOH, 0.2 mM disodium EDTA, pH 12; Truett et al. 2000). We placed the tubes in a Fisher Scientific™ Bead Mill 4 Homogenizer for 5 minutes at 5 m/s and centrifuged briefly to reveal un-popped eggs. We manually broke any un-popped eggs with sterile toothpicks. We then incubated the tubes at 95°C for 30 minutes, vortexing occasionally, and put them on ice for 3 minutes. Finally, we added 500 mL of HotSHOT neutralization buffer (40 mM Tris–HCl, pH 5; Truett et al. 2000) to each tube before storage at -20°C until the Polymerase Chain Reaction (PCR) step.

We PCR amplified a 226 base pair (bp) region of the mitochondrial cytochrome c oxidase I (COI) gene using primer set Mini_SH-E, with forward reads 5'-CACGACGTTGTAAAACGACACYAAICAYAAAGAYATIGGCAC-3', and reverse read 5'-GGATAACAATTTCACACAGGCTTATRTTRTTTATICGIGGRAAIGC-3' (Shokralla et al. 2015). Each 50-µl PCR contained final concentrations of 1x Apex NH_4 buffer, 1.5 mM Apex $MgCl_2$, 0.2 µM Apex dNTPs, 1 U Apex RedTaq (Genesee Scientific), 0.2 µM forward and reverse primers, 10 µg/µl bovine serum albumin (New England BioLabs Inc.), and 2 µl of target DNA (Burrows et al. 2019). The thermocycler conditions were as follows, 95°C for 5 min,

followed by 35 cycles of (94°C for 40 s, 46°C for 1 min, 72°C for 30 s) and 72°C for 5 min. We verified successful PCR amplification by running products on a 1.5% agarose gel stained with ethidium bromide. All samples were deemed successful and cleaned with the Zymo DNA Clean and Concentrator -25 Kit if the bands were bright, and the Zymo DNA Clean and Concentrator -5 Kit if the bands were faint. We quantified the amplified DNA using a Qubit™ dsDNA HS Assay Kit, normalized the samples to equal concentrations, and sent to Genewiz for next-generation Illumina sequencing using the Genewiz Amplicon-EZ pipeline and partial Illumina adapters, forward 5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3', reverse 5'-GACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3' to genetically identify the fish eggs present in each sample.

Sequence analysis

To process the raw sequence data and obtain filtered and trimmed amplicon sequence variants (ASV), we used the Divisive Amplicon Denoising Algorithm (DADA2) v1.12 package (Callahan et al. 2016) in the *R* statistical environment (R Core Team 2022). ASVs were first matched with species-level records in the Barcode of Life Database (BOLD; http://www.boldsystems.org/)(Sujeevan and Hebert 2007), then BLASTn comparison (Altschul et al. 1990) against the National Center for Biotechnology Information (NCBI) nucleotide database (https://blast.ncbi.nlm.nih.gov/Blast.cgi) if no match was made in BOLD. For some samples, more taxa were identified than the number of eggs present, signaling the presence of false positives and requiring the establishment of a threshold percentage of sequences required to consider a taxon present in a given sample (see Discussion). We applied a 2% threshold based on total number of eggs within each sample and number of taxa assigned. This empirically derived threshold is based on the principle that it would not be possible for a sample containing *n* number of eggs to have greater than *n* taxa present unless contamination was present. The final ASV table containing sequences that comprised >2% of the total sequence reads from any given sample is available in GRIIDC (Kerr et al 2023).

Quality control and data visualization

We identified sequences to the finest taxonomic resolution possible based on comparison to BOLD. In some cases, we could not distinguish between multiple potential identifications based on the sequenced portion of the COI gene. Therefore, to refine our identifications we referenced the geographic distribution of each taxon using published guides (McEachran and Fechhelm 1998, McEachran and Fechhelm 2005) and FishBase (Froese and Pauly 2019). We excluded any taxa not found in the Gulf of Mexico, such as those with Indo-Pacific or eastern Atlantic distributions. Twenty-one of the identifications were made at the species-level, while the remaining 16 identifications were to two or more closely related fishes; each distinct identification is referred to as a "taxon".

We also generated spatial heatmaps to visualize how metabarcoding-based identifications compared with the known habitat types occupied by each taxon (Fig. 1). Due to the qualitative

nature of metabarcoding identifications, we used a presence-based approach. We first categorized each of the 37 distinct taxa into one of eight habitat types, which included burrower, commensal, reef-associated, demersal, deep demersal, coastal pelagic, epipelagic, and mesopelagic, based on information from FishBase (Froese and Pauly 2019). We then quantified the number of distinct taxa in each habitat type at each sampling station. These presence-absence count data were then used to generate heatmaps in QGIS (QGIS Development Team 2022) where spatial weighting was set to 55,000 km (roughly equivalent to the distance between any two stations).

Results

We performed DNA metabarcoding on 4719 fish eggs from 49 samples, and obtained an average of 57,185 sequence reads per sample. The analyses presented here consider any taxa comprising ≥2% the sequence reads from a given sample to be "present" in the sample and any sequences comprising <2% the sequence reads from a given sample as "absent" since metabarcoding data are not quantitative due to methodological biases (see Discussion). We identified a total of 37 distinct taxa, with 1-11 distinct taxa per sample (Table 1; Supplemental Table 1). Twelve of the taxa (32%) were only detected in a single sample, eight taxa (22%) were detected in two samples, and the remainder were present in three or more samples. Tuna eggs (sequences could not be distinguished between *Auxis thazard/rochei*, *Euthynnus alletteratus*, and *Katsuwonus pelamis*) were identified in 26 samples, constituting the most widespread taxon in the dataset. Other taxa found at more than 10 sites included *Decapterus punctatus/tabl* (round/roughear scad), *Lutjanus griseus* (grey snapper), *Prionotus martis* (Gulf of Mexico barred searobin), *Pristipomoides aquilonaris* (wenchman), and *Xyrichtys sp.* (razorfish).

To validate the metabarcoding-based identification of fish eggs, we constructed spatial heatmaps based on presence-absence data (Fig. 1A-I), which were consistent with the expected distribution of the identified taxa. Specifically, reef-associated species were found broadly throughout the sampling area (Fig. 1D), corresponding with the patchy distribution of structured hard-bottom throughout the West Florida Shelf (Hine and Locker 2011). Coastal pelagic taxa were found inshore (Fig. 1G), epipelagic taxa were broadly distributed throughout the region (Fig. 1H), and mesopelagic taxa were found in deeper water farther offshore (Fig. 1I). Notably, we observed an apparent "hotspot" of demersal taxa (Fig. 1E) toward the southern end of the sampling region, which may indicate an area of interest for the management of recreationally or commercially important demersal fishes.

Discussion

DNA barcoding has gained popularity for identifying fish eggs; however, the vast majority of studies analyze individual fish eggs. Processing individual fish eggs yields quantitative data; that is, we directly determine the exact proportion of the collected fish egg community comprised by each taxon. Information on the abundance of planktonic fish eggs from

each taxon is valuable for estimating the biomass of parent fish stocks using the daily egg production method (Stratoudakis et al. 2006, Burrows et al. 2019). However, the number of eggs and sampling stations that can be processed with this method is limited by financial and labor resources. Metabarcoding of DNA extracted from all the collected fish eggs present at a given sample could be an advantageous alternative since it is faster and less expensive (Cristescu 2014). We are only aware of two studies to date that have applied metabarcoding to examine the community composition of fish eggs, one in marine waters and one in freshwater (Duke and Burton 2020, Miranda-Chumacero et al. 2020). Here we add to this emerging field of study by assessing the potential of metabarcoding as an alternate method for long-term monitoring of fish egg community composition. The advantages and disadvantages of DNA barcoding individual eggs versus the metabarcoding method applied here are summarized in Table 2 and discussed below.

Compared to DNA barcoding of individual fish eggs, in which a single sequence is obtained from each egg, metabarcoding is not quantitative due to variable copy number of mitochondrial DNA in different taxa, different numbers of cells within eggs depending on developmental stage, chimeric sequences, and PCR amplification biases (Bik et al. 2012, Hatzenbuhler et al. 2017, Duke and Burton 2020). A ground-truthing study by Duke and Burton (2020) demonstrated that fish egg metabarcoding reliably detected taxa that comprised over six percent of a mock community and three percent of a natural community, with variable recovery of rarer community members. Despite some variability, Duke and Burton (2020) found a positive relationship between the proportion of reads from a given taxon and the proportion of eggs from that taxon in the mock communities. Therefore, although metabarcoding data cannot provide absolute taxon proportions, this technique can yield valuable information about abundant taxa, as well as rarer taxa above a given threshold.

Another potential flaw with DNA metabarcoding of fish eggs is the detection of false positives either due to environmental DNA (eDNA) stuck on the fish egg surfaces or contamination introduced during processing (Fritts et al. 2019, Duke and Burton 2020). Duke and Burton (2020) found that most false positives comprised a very small percentage of the sequences recovered from a given sample. These data and other studies of fish early life stages suggest setting a threshold proportion of sequences required for counting a taxon as present; however, there is no consensus on what that threshold value should be and it may need to be specific to each study area (Mariac et al. 2018, Duke and Burton 2020, Miranda-Chumacero et al. 2020). In the present study, we empirically derived a threshold based on the total number of eggs within each sample and the number of taxa assigned, based on the principle that it would not be possible for a sample containing *n* number of eggs to have greater than *n* taxa unless contamination was present. We found that setting a 2% threshold (i.e., sequences comprising less than 2% of the total sample reads were considered false positives and removed) ensured that the maximum number of taxa never exceeded the maximum number of eggs in a sample.

Unlike individual egg DNA barcoding where a single sequence is recovered from each egg and the results are quantitative, we found that the metabarcoding results varied depending on

the threshold applied, thus false positives (i.e., sequences above the threshold that were not derived from eggs) and false negatives (i.e., sequences below the threshold that were derived from eggs but comprised a percentage of the sequences below the cutoff) remain problematic and can have a major effect on reported spawning sites. For example, the 2% threshold used to analyze the data in this study resulted in an average of 3.67 taxa per sample (range 1-11). Applying a more conservative 5% threshold would have resulted in an average of 2.18 taxa per sample (range 1-6). Although 32 of the 37 taxa identified in this study would still be detected in at least one sample with the 5% threshold, five taxa (*Acanthocybium solandri*/*Scomberomorus cavalla*, *Engraulis eurystole*, *Haemulon plumieri*, *Scomberomorus maculatus/regalis*, *Synagrops bellus/spinosus*) would have been removed completely. Eggs from all five of these taxa have been previously recovered on the West Florida Shelf through individual egg barcoding (Keel et al. 2022, Kerr et al. 2022) so in this case, we believe that increasing the threshold would likely result in false negatives. This example demonstrates the large effect that small differences in the threshold can have on DNA metabarcoding results.

Since the inception of the Fish Barcode of Life (FISH-BOL) Initiative, the COI gene is widely used for genetic identification of fishes and this gene is capable of distinguishing between the majority of described fish species (Teletchea 2009, Ward et al. 2009). Numerous studies have demonstrated the advantage of using multiple genetic markers for metabarcoding (Evans et al. 2016, Sawaya et al. 2019, Duke and Burton 2020); however, databases tend to be more limited for other markers and may need to be supplemented for geographic regions of interest. One advantage of individual egg barcoding is the ability to return to DNA samples extracted from specific eggs to analyze population genetics or apply additional primer sets in the case where the conserved region of the COI gene used for barcoding cannot distinguish between certain species complexes. For example, Burrows et al. (2019) applied additional PCR primers to distinguish between the economically important species *Thunnus thynnus* and *Katsuwonus pelamis*, as well as between *Scomberomorus cavalla* and *Acanthocybium solandri*, to achieve a definitive identification. The ability to return to specific eggs to refine taxonomic uncertainties is lost in DNA metabarcoding, where the DNA from all eggs within a sample is combined. Finally, it is possible that the shorter sequence length used for Illumina sequencing in metabarcoding compared to Sanger sequencing used for individual egg barcoding might hinder the assignment of sequences to species level. We did not experience lower taxonomic resolution due to the shorter sequence length obtained in this study, which is consistent with other analyses that have shown the ability to reliably assign 140 bp reads with relatively high success rates (Shokralla et al. 2015, Kimmerling et al. 2018, Mariac et al. 2018).

PCR biases represent an important barrier to the feasibility of metabarcoding studies for quantitative analyses of fish early life stages (Lamb et al. 2019, Zinger et al. 2019). Efforts to make metabarcoding of fish larvae more quantitative have shown success, although these methods have not yet been applied to eggs and have only been examined in a limited number of studies and regions. Applying a different approach to quantify fish early life stages, Kimmerling et al. (2018) used high coverage metagenomic sequencing (sequencing of total DNA from a

given sample, without first applying PCR to enrich for the COI gene). Kimmerling et al. (2018) adjusted the sequence coverage for each sample to obtain ~20 COI-derived reads per larva, allowing samples with more larvae to be sequenced more deeply. These methods yielded quantitative results when normalized by the relative size of each larva in the sample, showing the promise of this technique (Kimmerling et al. 2018). However, this method requires a large amount of sequencing since the percentage of metagenomic sequence reads that belonged to the COI gene was extremely low (approximately 1 in every 18,000 sequences), which will present a substantial barrier for long-term monitoring efforts. With a newly developed Metabarcoding by Capture using a Single Probe (MCSP) method, Mariac et al. (2018) achieved ~6,000 times enrichment of COI sequences compared to an unenriched sample. By analyzing a mock community, the relative frequencies of sequences recovered from larval swarms in the Amazon basin with the MCSP method correlated extremely well with true frequencies derived from Sanger sequencing of individual fish larvae (Mariac et al. 2018). Since this method relies on hybridization instead of PCR amplification, MCSP is subject to fewer biases. However, it should be noted that MCSP still required the application of a threshold value for the minimum number of reads per taxon in order to count that taxon as present. Even with the application of an empirically defined threshold established through the analysis of mock communities, a small number of false positives were still encountered (Mariac et al. 2018).

Finally, although we cannot directly compare the metabarcoding results presented here to our prior surveys of individual fish eggs in the GOM, it is notable that the vast majority of the taxa identified here have also been recovered from this region in our prior work (Keel et al. 2022, Kerr et al. 2022). The only exceptions found with metabarcoding that we have not observed with our more spatially and temporally expansive individual egg barcoding were *Makaira nigricans* (Atlantic blue marlin), *Lepidopus altifrons* (crested scabbardfish), and *Chilomycterus sp*. (burrfish), all of which are known to occur in the GOM. We examined the data for these taxa to determine if they were present in very low abundances and thus were likely false positives; however, that was not the case. *Chilomycterus* sp. comprised ~87% of the sequences in one sample, *Lepidopus altifrons* comprised ~20% of the sequences in another sample, and *Makaira nigricans* was found in two samples, where it made up ~16% of the sequences of each sample. This suggests that eggs from these taxa were truly present in these samples, and demonstrates an advantage of metabarcoding; namely, by enabling the processing of samples from more stations, we can capture rarer or more episodic spawning events.

Conclusions

In this study, we assessed the performance of DNA metabarcoding to increase throughput and reduce financial and labor costs associated with a long-term fish egg monitoring program. A total of 37 taxa were identified from 49 stations on the West Florida Shelf. Egg identifications were consistent with prior species distributions observed from individual egg DNA barcoding and spatial heatmaps of eggs corresponded to known habitat types occupied by adults. The

317 increased throughput allowed by metabarcoding resulted in the identification of taxa not
318 previously detected in this region, possibly representing episodic spawning events. One
319 disadvantage of metabarcoding is that this method is not quantitative and requires the application
320 of a threshold proportion of sequences required to count a taxon as present. The choice of DNA
321 barcoding methods therefore depends on the goals of the study, and fisheries monitoring efforts
322 may benefit from a combination of the two approaches, with individual egg barcoding providing
323 quantitative information and metabarcoding expanding the number or geographic range of
324 samples that can be processed.
325
326

327 Acknowledgements

328

329 The authors thank the volunteers that collected plankton samples and the captain and
330 crew on the R/V Hogarth in August and September, 2019. Thanks to Claudia Baron-Aguilar,
331 Brianna Michaud, Vi Nguyen, Julie Vecchio, and Amy Wallace for help picking fish eggs from
332 the plankton samples, and to Natalie Sawaya for guidance with bioinformatics.
333
334

Tables

Table 1. Taxa comprising more than 2% of the sequences from any sample, habitat classification, and the number of stations where the taxa were identified.

Table 2. Comparison of individual egg DNA barcoding versus metabarcoding.

Figures

Figure 1. West Florida Shelf study region. (A) Station locations were defined a priori, and heatmaps were based on the number of taxa identified at each station for each habitat type: (B) Burrower, (C) Commensal, (D) Reef-Associated, (E) Demersal, (F) Deep Demersal, (G) Coastal Pelagic, (H) Epipelagic, and (I) Mesopelagic.

Supplemental Files

Supplemental Data S1. Detailed table of taxa identified in each sample, including scientific name, common name, and habitat type. Metadata for each sample include the station number, latitude and longitude, maximum water depth, date collected, and number of eggs processed.

References

Ahern, A. L. M., J. Gómez-Gutiérrez, O. Aburto-Oropeza, R. J. Saldierna-Martínez, A. F. Johnson, A. E. Harada, A. R. Sánchez-Uvera, B. Erisman, D. I. C. Arvizú and R. S. Burton (2018). "DNA sequencing of fish eggs and larvae reveals high species diversity and seasonal changes in spawning activity in the southeastern Gulf of California." Marine Ecology Progress Series **592**: 159-179.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). "Basic local alignment search tool." Journal of Molecular Biology **215**: 403-410.

Becker, R., N. G. Sales, G. Santos, G. Santos and D. Carvalho (2015). "DNA barcoding and morphological identification of neotropical ichthyoplankton from the Upper Paraná and São Francisco." Journal of Fish Biology **87**: 159-168.

Bik, H. M., D. L. Porazinska, S. Creer, J. G. Caporaso, R. Knight and W. K. Thomas (2012). "Sequencing our way towards understanding global eukaryotic biodiversity." Trends in Ecology and Evolution **27**: 233.

Burghart, S., L. Van Woudenberg, C. Daniels, S. Myers, E. Peebles and M. Breitbart (2014). "Disparity between plankton fish egg and larval communities as indicated by DNA barcoding." Marine Ecology Progress Series **503**: 195-204.

Burrows, M., J. Browning, M. Breitbart, S. Murawski and E. Peebles (2019). "DNA barcoding reveals clear delineation between spawning sites for neritic versus oceanic fishes in the Gulf of Mexico." Fisheries Oceanography **28**: 228-239.

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

Cowen, R. K. and S. Sponaugle (2009). "Larval dispersal and marine population connectivity." Annual Review of Marine Science **1**: 443-466.

Cristescu, M. E. (2014). "From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity." Trends in Ecology & Evolution **29**: 566-571.

Duke, E. M. and R. S. Burton (2020). "Efficacy of metabarcoding for identification of fish eggs evaluated with mock communities." Ecology and Evolution **10**: 3463-3476.

Duke, E. M., A. E. Harada and R. S. Burton (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.

Evans, N. T., B. P. Olds, M. A. Renshaw, C. R. Turner, Y. Li, C. L. Jerde, A. R. Mahon, M. E. Pfrender, G. A. Lamberti and D. M. Lodge (2016). "Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding." Molecular Ecology Resources **16**: 29-41.

Frantine-Silva, W., S. Sofia, M. Orsi and F. Almeida (2015). "DNA barcoding of freshwater ichthyoplankton in the Neotropics as a tool for ecological monitoring." Molecular Ecology Resources **15**: 1226-1237.

Fritts, A. K., B. C. Knights, J. H. Larson, J. J. Amberg, C. M. Merkes, T. Tajjioui, S. E. Butler, M. J. Diana, D. H. Wahl and M. J. Weber (2019). "Development of a quantitative PCR method for screening ichthyoplankton samples for bigheaded carps." Biological Invasions **21**: 1143-1153.

Froese, R. and D. Pauly (2022). FishBase, World Wide Web electronic publication. www.fishbase.org.

Harada, A. E., E. A. Lindgren, M. C. Hermsmeier, P. A. Rogowski, E. Terrill and R. S. Burton (2015). "Monitoring spawning activity in a southern California marine protected area using molecular identification of fish eggs." PLoS One **10**: e0134647.

Hatzenbuhler, C., J. R. Kelly, J. Martinson, S. Okum and E. Pilgrim (2017). "Sensitivity and accuracy of high-throughput metabarcoding methods for early detection of invasive fish species." Scientific Reports **7**: 46393.

Hine, A. C. and S. D. Locker (2011). "Florida Gulf of Mexico contitential shelf great contrasts and significant transitions". In: Gulf of Mexico Origin, Waters, and Biota: Volume 3, Geology. Edited by N. A. Buster and C. W. Holmes. Texas A&M University Press.

Hofmann, T., T. Knebelsberger, M. Kloppmann, J. Ulleweit and M. Raupach (2017). "Egg identification of three economical important fish species using DNA barcoding in comparison to a morphological determination." Journal of Applied Ichthyology **33**: 925-932.

Hou, G., Y. Chen, J. Wang, C. Pan, J. Lin, B. Feng and H. Zhang (2022). "Molecular identification of species diversity using pelagic fish eggs in spring and late autumn-winter in the Eastern Beibu Gulf, China." Frontiers in Marine Science **8**: 806208.

Hou, G., J. Wang, Z. Chen, J. Zhou, W. Huang and H. Zhang (2020). "Molecular and morphological identification and seasonal distribution of eggs of four decapterus fish species in the northern South China Sea: A key to conservation of spawning ground." Frontiers in Marine Science **7**: 590564.

Kawakami, T., J. Aoyama and K. Tsukamoto (2010). "Morphology of pelagic fish eggs identified using mitochondrial DNA and their distribution in waters west of the Mariana Islands." Environmental Biology of Fishes **87**: 221-235.

Keel, K., M. Kerr, E. Peebles, C. Stallings, G. Zapfe and M. Breitbart (2022). Spawning habitat and early-life linkages to fisheries (SHELF) - 2013 and 2014 continuous underway fish egg sampler (CUFES) fish egg sequences. GRIIDC. doi: 10.7266/j7qzhk64.

Kerr, M., J. Browning, E. M. Bønnelycke, Y. Zhang, C. Hu, M. Armenteros, S. Murawski, E. Peebles and M. Breitbart (2020). "DNA barcoding of fish eggs collected off northwestern Cuba and across the Florida Straits demonstrates egg transport by mesoscale eddies." Fisheries Oceanography **29**(4): 340-348.

Kerr, M., G. Koziol, E. Peebles, C. Stallings, G. Zapfe and M. Breitbart (2022). Spawning habitat and early-life linkages to fisheries (SHELF) - 2019 continuous underway fish egg sampler (CUFES) fish egg sequences. GRIIDC. doi: 10.7266/dmp83r1m.

Kerr, M., G. Koziol, E. Peebles, C. Stallings and M. Breitbart (2023). 2019 metabarcoding fish eggs on the West Florida Shelf. GRIIDC. doi: 10.7266/pvp5zc1.

Kimmerling, N., O. Zuqert, G. Amitai, T. Gurevich, R. Armoza-Zvuloni, I. Kolesnikov, I. Berenshtein, S. Melamed, S. Gilad and S. Benjamin (2018). "Quantitative species-level ecology of reef fish larvae via metabarcoding." Nature Ecology & Evolution **2**(2): 306.

Lamb, P. D., E. Hunter, J. K. Pinnegar, S. Creer, R. G. Davies and M. I. Taylor (2019). "How quantitative is metabarcoding: A meta-analytical approach." Molecular Ecology **28**(2): 420-430.

Lelièvre, S., M. Jérôme, G. E. Maes, S. Vaz, S. Calaivany and V. Verrez-Bagnis (2012). "Integrating molecular identification of pelagic eggs with geostatistical mapping to

improve the delineation of North Sea fish spawning grounds." Marine Ecology Progress Series **445**: 161-172.
Lewis, L. A., D. E. Richardson, E. V. Zakharov and R. Hanner (2016). "Integrating DNA barcoding of fish eggs into ichthyoplankton monitoring programs." Fishery Bulletin **114**(2): 153-165.
Leyva-Cruz, E., L. Vásquez-Yeomans, L. Carrillo and M. Valdez-Moreno (2016). "Identifying pelagic fish eggs in the southeast Yucatan Peninsula using DNA barcodes." Genome **59**(12): 1117-1129.
Lin, H.-Y., M.-Y. Chiu, Y.-M. Shih, I.-S. Chen, M.-A. Lee and K.-T. Shao (2016). "Species composition and assemblages of ichthyoplankton during summer in the East China Sea." Continental Shelf Research **126**: 64-78.
Mariac, C., Y. Vigouroux, F. Duponchelle, C. García-Dávila, J. Núñez, E. Desmarais and J.-F. Renno (2018). "Metabarcoding by capture using a single COI probe (MCSP) to identify and quantify fish species in ichthyoplankton swarms." PLoS One **13**(9): e0202976.
Mateos-Rivera, A., R. Skern-Mauritzen, G. Dahle, S. Sundby, B. Mozfar, A. Thorsen, H. Wehde and B. A. Krafft (2020). "Comparison of visual and molecular taxonomic methods to identify ichthyoplankton in the North Sea." Limnology and Oceanography: Methods **18**(10): 599-605.
McEachran, J. D. and J. D. Fechhelm (1998). Fishes of the Gulf of Mexico, Volume 1: Myxiniformes to Gasterosteiformes. Austin, University of Texas Press.
McEachran, J. D. and J. D. Fechhelm (2005). Fishes of the Gulf of Mexico, Volume 2: Scorpaeniformes to Tetraodontiformes. Austin, University of Texas Press.
Miranda-Chumacero, G., C. Mariac, F. Duponchelle, L. Painter, R. Wallace, G. Cochonneau, J. Molina-Rodriguez, C. Garcia-Davila and J.-F. Renno (2020). "Threatened fish spawning area revealed by specific metabarcoding identification of eggs and larvae in the Beni River, upper Amazon." Global Ecology and Conservation **24**: e01309.
QGIS Development Team (2022). QGIS Geographic Information System, QGIS Assocation.
Saitoh, K., S. Uehara and T. Tega (2009). "Genetic identification of fish eggs collected in Sendai Bay and off Johban, Japan." Ichthyological Research **56**: 200-203.
Sawaya, N. A., A. Djurhuus, C. J. Closek, M. Hepner, E. Olesin, L. Visser, C. Kelble, K. Hubbard and M. Breitbart (2019). "Assessing eukaryotic biodiversity in the Florida Keys National Marine Sanctuary through environmental DNA metabarcoding." Ecology and Evolution **9**: 1029-1040.
Shao, K.-T., K.-C. Chen and J.-H. Wu (2002). "Identification of marine fish eggs in Taiwan using light microscopy, scanning electric microscopy and mtDNA sequencing." Marine and Freshwater Research **53**: 355-365.
Shokralla, S., R. S. Hellberg, S. M. Handy, I. King and M. Hajibabaei (2015). "A DNA mini-barcoding system for authentication of processed fish products." Scientific Reports **5**: 15894.
Stratoudakis, Y., M. Bernal, K. Ganias and A. Uriarte (2006). "The daily egg production method: recent advances, current applications and future challenges." Fish and Fisheries **7**: 35-57.
Sujeevan, R. and P. Hebert (2007). "BOLD: the barcode of life data system." Molecular Ecology Notes **7**: 355-364.
Teletchea, F. (2009). "Molecular identification methods of fish species: reassessment and possible applications." Reviews in Fish Biology and Fisheries **19**: 265-293.

492 Truett, G., P. Heeger, R. Mynatt, A. Truett, J. Walker and M. Warman (2000). "Preparation of
493 PCR-quality mouse genomic DNA with hot sodium hydroxide and Tris (HotSHOT)."
494 Biotechniques **29**: 52-54.
495 Ward, R. D., R. Hanner and P. D. Hebert (2009). "The campaign to DNA barcode all fishes,
496 FISH-BOL." Journal of Fish Biology **74**: 329-356.
497 Zinger, L., A. Bonin, I. G. Alsos, M. Bálint, H. Bik, F. Boyer, A. A. Chariton, S. Creer, E.
498 Coissac and B. E. Deagle (2019). "DNA metabarcoding—Need for robust experimental
499 designs to draw sound ecological conclusions." Molecular Ecology **28:** 1857-1862.

Table 1(on next page)

Taxa comprising more than 2% of the sequences from any sample, habitat classification, and the number of stations where the taxa were identified.

Table 1: Taxa comprising more than 2% of the sequences from any sample, habitat classification, and the number of stations where the taxa were identified.

Scientific Name	Common Name	Habitat	# Stations
Acanthocybium solandri, *Scomberomorus cavalla*	Wahoo/King Mackerel	Epipelagic	3
Acanthostracion quadricornis	Scrawled Cowfish	Reef	6
Auxis thazard/rochei, *Euthynnus alletteratus*, *Katsuwonus pelamis*	Bullet/Frigate Tuna, Little Tunny, Skipjack	Epipelagic	26
Brama dussumieri/caribbea	Lesser Bream/Carribean Pomfret	Deep demersal	4
Callechelys muraena	Blotched Snake Eel	Burrower	1
Caranx crysos	Blue Runner	Coastal pelagic	5
Chaetodipterus faber	Atlantic Spadefish	Reef	1
Chilomycterus schoepfii/antillarum	Striped/Web Burrfish	Reef	1
Coryphaena hippurus	Common Dolphinfish	Epipelagic	2
Cyclopsetta fimbriata	Spotfin Flounder	Demersal	3
Decapterus punctatus/tabl	Round/Roughear Scad	Coastal pelagic	16
Diplogrammus pauciradiatus	Spotted Dragonet	Reef	1
Diplospinus multistriatus/Eustomias polyaster	Striped escolar/Dragonfish	Mesopelagic	5
Echeneis naucrates/neucratoides, *Remora remora/osteochir*	Live/Whitefin Sharksucker, Marlin/Sharksucker	Commensal	8
Engraulis eurystole	European/Silver anchovy	Coastal pelagic	1
Gordiichthys irretitus	Horsehair Eel	Burrower	1
Haemulon plumieri	White Grunt	Reef	1
Lepidopus altifrons	Crested Scabbardfish	Mesopelagic	1
Lutjanus campechanus	Red Snapper	Reef	2
Lutjanus griseus	Grey Snapper	Reef	11
Makaira nigricans	Blue Marlin	Epipelagic	2
Oxyporhamphus similis/micropterus	Halfbeaks	Epipelagic	4
Prionotus martis	Gulf of Mexico Barred Searobin	Demersal	11
Prionotus ophryas/scitulus	Bandtail/Leopard Searobin	Demersal	2
Prionotus roseus	Bluespotted Searobin	Demersal	2
Prionotus rubio/tribulus	Blackwing/Bighead Searobin	Demersal	8
Pristipomoides aquilonaris	Wenchman	Reef	18

Prognichthys occidentalis	Bluntnose Flyingfish	Epipelagic	2
Rachycentron canadum	Cobia	Reef	1
Rhomboplites aurorubens	Vermilion Snapper	Reef	2
Saurida normani/*brasiliensis*	Shortjaw/Brazilian Lizardfish	Demersal	2
Scomberomorus maculatus/*regalis*	Atlantic Spanish Mackerel/Cero	Coastal pelagic	1
Synagrops bellus/*spinosus*	Blackmouth/Keelcheek Bass	Mesopelagic	1
Synodus intermedius/*foetens*/ *macrostigmus*/sp	Sand Diver/Inshore/ Largespot/Lizardfish sp.	Demersal	5
Thunnus atlanticus/*albacares*/sp	Blackfin/Yellowfin/Tuna sp.	Epipelagic	8
Trachinocephalus myops	Snakefish	Demersal	1
Xyrichtys novacula/sp	Pearly Razorfish/Razorfish sp.	Burrower	11

Table 2(on next page)

Comparison of individual egg DNA barcoding versus metabarcoding.

Table 2: Comparison of individual egg DNA barcoding versus metabarcoding.

Parameter	Individual eggs	Metabarcoding
Cost	$5.15 per egg $494.40 per site[a]	$0.78 per egg[a] $64.82 per site
Sequencing platform	Sanger	Illumina
Average sequence length	500 base pairs	200 base pairs
Quantitative	Yes	No
Ability to return to individual eggs with additional primers	Yes	No
Prevalence of false positives/negatives	Low/None	Frequent; dependent on the application of a threshold

[a] Cost calculated based on 96 eggs per site.

Figure 1

West Florida Shelf study region and heatmap of identified taxa

(A) Station locations were defined a priori, and heatmaps were based on the number of taxa identified at each station for each habitat type: (B) Burrower, (C) Commensal, (D) Reef-Associated, (E) Demersal, (F) Deep Demersal, (G) Coastal Pelagic, (H) Epipelagic, and (I) Mesopelagic.

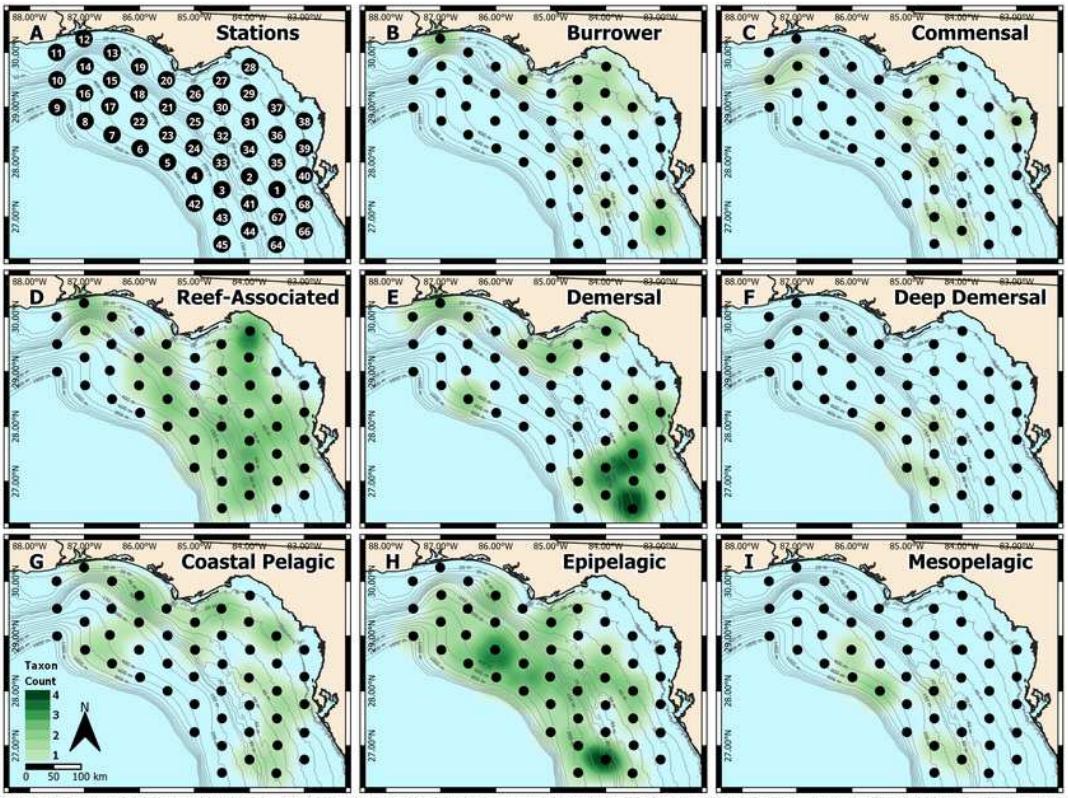

Figure 1: West Florida Shelf study region. Station locations were defined a priori (A), and heatmaps were based on the number of taxa identified at each station for each habitat type: Burrower (B), Commensal (C), Reef-Associated (D), Demersal (E), Deep Demersal (F), Coastal Pelagic (G), Epipelagic (H), and Mesopelagic (I).