Developing a DNA Barcode Database of Gulf of Mexico Fishes for Biodiversity Research

Tanner Bahm and Kyle R. Piller
Southeastern Louisiana University, Dept. of Biological Sciences
January 2022

Abstract

The overall goal of this project was to fill in data gaps in the DNA barcoding reference databases by gathering sequence data from inshore Gulf of Mexico fishes. Accurate identification of species represents an important first step in environmental DNA studies. DNA sequence data, 12S mitochondrial DNA, was gathered from 59 individuals for which 12S sequence data was not previously available. These DNA sequences will be added to Genbank and serve as baseline data for future and on-going projects in the region, thereby making the data available to researchers across the globe.

Introduction

Coastal regions of the northern Gulf of Mexico are some of the most productive ecosystems in the world (Nixon 1980, Houde and Rutherford 1993). In particular, estuarine habitats are species rich, spatially complex, and are regularly used by estuarine and marine fishes at one or more life-history stages (O'Connell et al. 2004). Despite the high levels of productivity, many estuaries and coastal regions in southeastern Louisiana possesses little natural "hard structure" and are dominated by soft-bottomed sediments. As a result, an artificial reef program was instituted to provide additional essential fish habitat through the deployment of clusters of concrete reef balls or the distribution of crushed limestone/concrete. The use of artificial habitats follows the Field of Dream approach ("*if you build it they will come*"). However, elsewhere, these types of structures have proven to be effective as a fish attractant tool and have enhanced fisheries in marine and estuarine environments. Overall, they can provide food resources by supplying hard substrates for colonizing invertebrates, while simultaneously providing structures for shelter and reproduction for many fish species.

At the present time, other than a few recreationally fished species, it is unclear as to which fish species are utilizing the artificial reefs in southeast Louisiana and which reef material attracts the greatest number of species. This is due, in part, to the difficulty in monitoring artificial reefs. The high turbidity in the region makes it challenging (or nearly impossible) to assess fish communities of artificial reefs. Elsewhere, artificial reefs are typically visually monitored via SCUBA or ROVs, which is clearly not feasible in the northern Gulf of Mexico. Understanding fish diversity on these reefs is the first step to addressing and managing a sustainable fishery.

Environmental DNA (eDNA) is a relatively new approach that is being used to monitor the presence of target species in a variety of habitats, including artificial reefs. Environmental DNA is a non-invasive approach used to monitor the occurrence of species in a given area

without physically encountering a particular species (Jerde et al. 2011). Environmental DNA is any organic bio-product of an organism in the environment including shed scales, mucous, feces, urine, etc. (Thomsen and Willerslev 2015). It has proven to be a very valuable approach to monitor populations of aquatic organisms, including rare or uncommon species, which are often undetected during traditional sampling.

Analytically, the eDNA approach utilizes a bioinformatics pipeline, known at Mitofish, that processes the recovered Illumina sequence data, and then uses multiple reference sequence databases (i.e. Genbank) to assign unknown DNA sequence reads to a particular fish species (metabarcoding). More than 1,440 species of fish have been documented to occur in the Gulf of Mexico, but only a small proportion of the 12S sequences for these species are represented in Genbank and other publically available reference databases, thereby limiting the species level resolution of the eDNA recovered on artificial reefs. It is estimated that approximately 250 species occupy inshore areas of the Gulf of Mexico, but only about half of these have 12S sequences represented in the Genbank database. Database limitations also have been identified in other eDNA metabarcoding studies, but massive barcoding endeavors in other areas are underway in an attempt to augment these databased. Therefore, the goal of this project was to fill data gaps in the reference databases by sequencing the 12S marker for as many inshore Gulf of Mexico species as possible. The recovered sequence data will be beneficial to eDNA studies of fish communities in the Gulf of Mexico and beyond.

Methods

A comprehensive list of inshore species known to occur in the Gulf of Mexico was developed from the literature. Once the species list was developed, we searched Genbank to check the availability of 12S sequences that cover the barcode region amplified by the primers used in Miya et al. (2015). Then, with this information in hand, we searched a variety of publically available museum databases, the tissue collection at Southeastern's Vertebrate Museum, as well as contacting various colleagues to determine the availability of tissue samples.

Ultimately, tissue samples were obtained from colleagues at Tulane University, Southeastern Louisiana University's Vertebrate Collection, NOAA-Pascagoula, and our own sampling. DNA was extracted from tissue samples using a DNeasy tissue kit (Qiagen) and amplified using the MiFish-F primer (Miya et al. 2015) and the AcMDB07R (Bylemans et al. 2018) under the cycling parameters provided in **Table 1**. PCR amplicons were visualized on a 0.8% agarose gel and submitted to Genewiz for Sanger sequencing. The recovered sequences were aligned to a full reference 12S sequence from Genbank. These data will be submitted to Genbank. Once these samples are available on Genbank, then the Mitofish pipeline can incorporate this data into current and future eDNA studies.

Table 2. PCR cycling parameters

Initial Denaturation	95°C for 2 minutes	x 1 cycle
Denaturation	95°C for 35 secs	
Extension	57°C for 35 seconds	x 35 cycle
Annealing	72°C for 45 seconds	
Final Annealing	72°C for 10 minutes	x 1 cycle

Results/Discussion

A total of 59 tissue samples from 59 species, across 50 genera and 30 families were obtained and 12S mitochondrial sequence data was previously unavailable for these species. The recovered sequences were trimmed to a 475 fragment that spans the barcode region of Miya et al. (2015) (**Fig. 1**). Across all samples the base composition for the 475bp fragment consisted of T (21%), C (27.8%), A (29.6%), and G (21.7%).

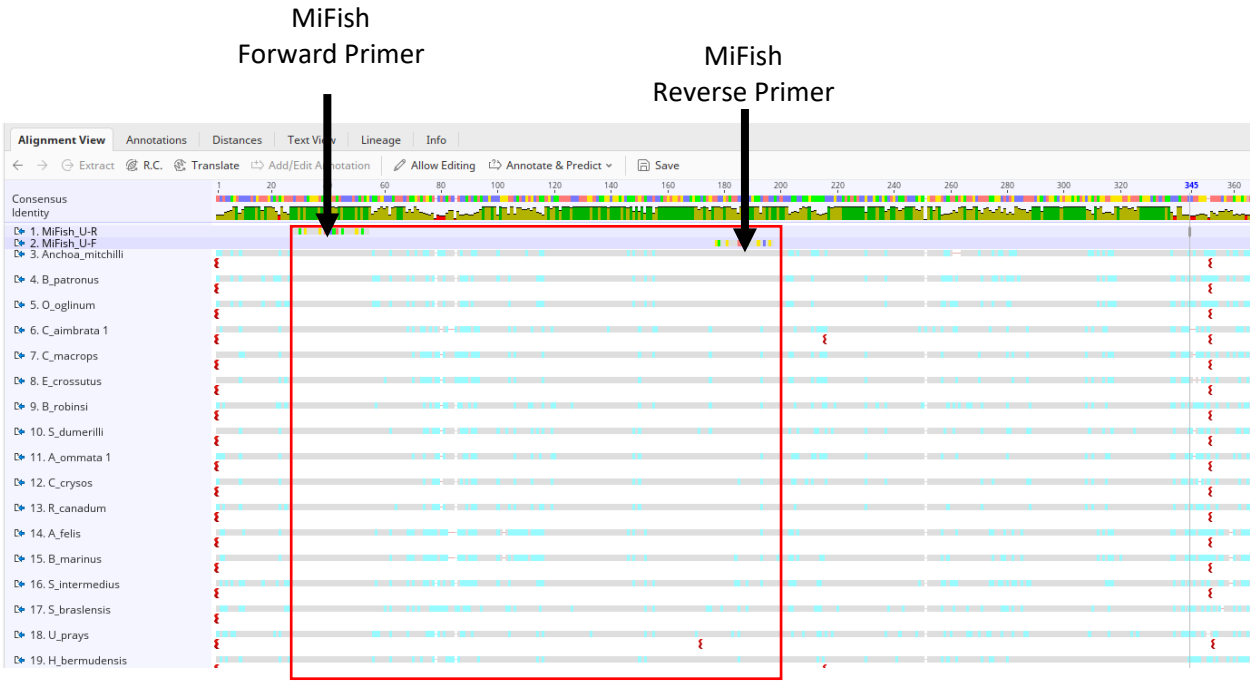

Fig. 1. DNA sequence alignment of 12S (mtDNA) for a select group of fishes from the Gulf of Mexico. The highlighted area within the red box corresponds to the barcode region amplified by the Miya et al. (2015) primers used in eDNA metabarcode studies.

Pairwise differences (uncorrected p-distance) between species ranged from 5.40-31.70% (**Table 2**). Pairwise comparisons between families ranged from 10.1% (Poeciliidae vs. Cyprinodontidae) to 27.8% (Ariidae vs. Mullidae). Within families, the average sequence variation ranged from 5.8 (Haemulidae)-20.7% (Serranidae).

Table 2. Genetic divergence (uncorrected-p distance) at different taxonomic levels.

Group	N	Mean Difference (%)	Min (%)	Max (%)
Species	59	19.9	5.4	31.7
Family	50	20.2	10.1	27.8

Discussion

This UROP proposed project produced more than 50 new 12S sequences that will be uploaded to Genbank. Inclusion of these samples to this federal repository will assure that they are available to those interested in genetics and eDNA studies of Gulf of Mexico fishes. In particular, the justification for gathering this sequence data is to augment the reference database for Gulf of Mexico fishes for current and future eDNA studies.

Relative to other fishery sampling approaches, eDNA monitoring is in its infancy with protocol modifications and technique improvements constantly occurring. The accurate genetic identification of a taxon requires the existence of a comprehensive reference database. In the case of the Gulf of Mexico fishes, the lack of a comprehensive 12S mtDNA reference database is a substantial issue, as the existence of incomplete reference databases can lead to underestimates of biodiversity in eDNA studies since recovered DNA sequences are assigned to known reference sequences. The data collected in this study, and hopefully many more in the future, will help eliminate this known issue. The upside of augmenting reference sequences from any study region is that once databases are updated the original metabarcode sequence data generated in any study could be resubmitted to the analytical pipeline. The results can be re-analyzed to recover a much more complete and comprehensive picture of an eDNA based fish assemblage using the more complete database.

Literature Cited

Bylemans J, D. M. Gleeson, C. M. Hardy, and E. Furlan. 2018. Toward an ecoregion scale evaluation of eDNA metabarcoding primers: A case study for the freshwater fish biodiversity of the Murray–Darling Basin (Australia). Ecol Evol. 8:8697-8712. 10.1002/ece3.4387

Houde, E.D., and E. S. Rutherford. 1993. Recent trends in estuarine fisheries: Predictions of fish production and yield. Estuaries 16:161-176.

Jerde, C. L., A. R. Mahon, W. L. Chadderton, and D. M. Lodge. 2011. Sight-unseen detection of rare aquatic species using environmental DNA. Conserv. Lett. 4:150–157

Miya, M., Y. Sato, T. Fukunaga, T. Sado, J. Y. Poulsen, Sato, etc. 2015. MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: Detection of more than 230 subtropical marine species. Royal Society Open Science, 2, 150088 10.1098/rsos.150088

O'Connell, M. T., R. C. Cashner and C. S. Schieble. 2004. Fish assemblage stability over fifty years in the Lake Pontchartrain estuary; Comparisons among habitats using canonical correspondence analysis. Estuaries 27: 807-817.

Nixon, S. W. 1980. Between coastal marshes and coastal waters-a review of twenty years of speculation and research on the role of salt marshes in estuarine productivity and water chemistry, p. 437-525. In P. Hamilton and K. B. MacDonald (eds.), Estuarine and Wetland Processes. Plenum, New York.

Thomsen, P. F., J. Kielgast, L. L. Iversen, P. R. Møller, M. Rasmussen. 2012. Detection of a diverse marine fish fauna using environmental DNA from seawater samples. PLoS ONE 7: e41732. doi:10.1371/journal.pone.0041732.

Thomsen, P. F. and E. Willerslev. 2015. Environmental DNA-An emerging tool in conservation for monitoring past and present biodiversity. Biological Conservation 183:4-18.