Forensic DNA genotyping of pelagic marine fish larvae sampled across the Western North Atlantic Continental Shelf

An essay submitted in partial fulfillment of

the requirements for graduation from the

Honors College at the College of Charleston

with a Bachelor of Science in

Marine Biology

Amanda M. Cole

May 2013

Advisor: Andrew M. Shedlock Ph.D.

Forensic DNA genotyping of pelagic marine fish larvae sampled across the Western North Atlantic Continental Shelf

Amanda M. Cole and Andrew M. Shedlock, Department of Biology

Abstract

This project generated an archive of purified genomic DNA from 300 pelagic marine fish larvae collected through the CofC *Transects* program from November 2003-November 2005 at 20 oceanic sampling stations spanning the continental shelf of South Carolina. High phenotypic plasticity of open-ocean larval organisms prohibits their reliable identification by conventional morphological character analysis. Therefore PCR amplification and DNA sequencing of Cytochrome Oxidase subunit 1 (CO1), a mitochondrial protein encoding gene, and the non-coding mitochondrial D-loop was used to forensically genotype larvae at two distinct loci. Of the 300 samples processed, over 100 yielded clean target sequences. PCR products were analyzed by agarose gel electrophoresis prior to sequencing, and online BLAST-based informatics of genomic data revealed taxonomic diversity of more than 25 distinct families of fishes. These successful genotyping results are being integrated with physical oceanography and biogeography to help bridge a large gap in our ability to model juvenile recruitment and community dynamics of ecologically important pelagic fish populations.

Introduction

Larval fish are often morphologically ambiguous in early stages, thus making them challenging to identify, even by a trained specialist. High phenotypic plasticity prohibits identification by morphological analysis. Their features are subject to dramatic changes as they develop, and the larvae often appear drastically different from the adult fishes. This is especially apparent in leptocephalus larvae, which are present in many diverse species of eels, bonefish, tarpon and other pelagic fishes. These larvae all have elongated, laterally compressed, transparent bodies (Richardson & Cowen 2004). There were many examples of leptocephalus larvae within the samples, and it was interesting to interrogate from which adult taxa they had been spawned. The only reliable way to identify these and other types of pelagic larval fish is through genetic analysis

This identification was done by using DNA sequence analysis of two genes undergoing different profiles of molecular evolution: Cytochrome Oxidase subunit 1 (CO1), a mitochondrial

protein encoding gene; and the D-Loop (also known as the Control Region or Displacement loop) of the mitochondrial genome. CO1 has come to be known as the "barcoding" gene because its nucleotide substitution profiles at specific codon positions are highly divergent among species (Hebert et al 2003). D-Loop is also highly variable among species because it does not experience purifying selection against amino acid replacements (as in coding sequences), and has been valuable in determining the evolutionary history of closely related vertebrates due to its high genetic diversity both among and within species (Pegg et al 2006). Taken together these two genes provide complementary levels of resolution across a broad range of molecular divergence times as measured by comparative DNA sequence analysis. Using the D-Loop in addition to CO1 adds significant discriminatory power to the present genotyping investigation. Other studies have been attempted using only CO1, but when high accuracy of DNA barcoding is needed, the single-locus approach can be problematic (Vences et al 2005).

The main objective of this project is to accurately genotype and forensically determine the taxonomic identification of a large and diverse sample of unknown pelagic marine fish larvae. By identifying these samples, a better understanding of which species of pelagic fishes inhabit the coastal waters can be attained. The samples used in this project were collected in 2003 through 2005 by the *Transects* program through the College of Charleston (Tarpey 2010). The larvae were collected from 20 sampling stations throughout the continental shelf of South Carolina from depths of 20-400m. A wide range of species were collected. Initial broad morphological evaluation identified filefish, porcupine fish, sea horses, and many other broad taxonomic groups present; however, it was not possible to use morphology to positively identify larvae at any specific taxonomic levels. From the several hundred individual larvae present, a

wide morphological range was chosen in order to maximize the most diverse taxonomic and oceanographic results.

The present project is significant for a number of reasons. The first is that the work greatly strengthens the scientific value of the *Transects* program at the College of Charleston. This program ran from 2003 to 2010 and was geared toward teaching undergraduate students the basics of biological and geological sampling at sea (oceanica.cofc.edu/ Transects). Our project was an opportunity to continue the analysis of samples collected, so that the academic benefits of the program can live on in our work. Given the nature of the survey, many commercially and ecologically important pelagic species were expected to be present in the samples. For this project, genomic DNA from 300 samples was extracted, which gave us an opportunity to learn about the broad community structure and patterns of larval dispersal present over the continental shelf off the coast of South Carolina.

The present project is also relevant and important to other professional researchers such as ichthyologists and developmental biologists, because it furthers the DNA genotyping efforts already taking place, and supports a new effective way for identifying ambiguous larvae (Savolainen et al 2005). The analysis revealed common types of larvae present over the continental shelf off of South Carolina at multiple depths. This information can be significant to many different fields such as fisheries management, environmental management, and climate change research. By identifying many of these larvae collected, patterns of larval recruitment and community diversity in the waters surrounding South Carolina can be analyzed. The species present are all important to the health of the pelagic community, so determining their population range and dispersal is valuable in predicting or detecting environmental changes due to anthropogenic pressures such as pollution, climate change and ocean acidification.

Contamination, temperature shifts and habitat modification can greatly change the diversity of larval fishes in affected areas (McKinley et al 2011). Ocean acidification may also lead to smaller size and developmental delays in some species of pelagic marine fishes (Bignami et al 2013). These anthropogenic pressures and others could greatly negatively affect the population diversity of larvae as well as hinder their individual growth and development into healthy adults.

A large portion of the samples are commercially important species, so this work can be a baseline study in determining population density of species that are of economic interest. The results of this project can be directly applied to solving potential ecological problems off the coast of the Carolinas, which is something to which this study hopes to contribute. This investigation can directly inform future studies on climate change, pollution, and the effects of overfishing by providing important baseline data on species richness and abundance in the Western Atlantic.

Materials and Methods

The larvae samples were collected by Neuston and Bongo nets during the College of Charleston *Transects* program November 2003-November 2005, which allowed for obtaining a wide range of sizes and species of larvae. The samples were received from Dr. Gorka Sancho, who had them in long-term storage in ethanol at the Grice Marine Laboratory. 300 individuals were chosen that spanned a wide range of species so the analysis would have a broad scope.

After choosing the desired larvae, the Qiagen DNeasy Blood and Tissue Kit was used to isolate the DNA (www.qiagen.com). The procedure for this kit involves lysing the tissue, binding the DNA to a membrane in a spin-column, and washing it and eluting it with buffers

included in the kit. It is a fast, convenient, non-toxic method for collecting a high yield of DNA, even in small samples such as the developing larvae.

The next step in the process is taking a small amount of the DNA collected and doing a polymerase chain reaction (PCR) to replicate and get a large amount of DNA for sequencing (Palumbi et al 2002). The amplification of a particular region of DNA is executed by specific primers that bind to nucleotide sequences flanking the desired target region and cause DNA polymerase to build a copy. The amount of DNA grows exponentially because each new copy is used as a template for replication (Kocher et al 1989).

For these reactions, universal primers from Integrated DNA Technologies were used (idtdna.com/site). A number of so-called "universal" primers were developed by other studies (Kocher et al, 1989; Palumbi et al, 2002; Ward et al 2005), and the sequences of primers used with recent success are taken from Tarpey (2010) for CO1 and from Gomes et al (1997) for D-Loop:

UniCO1-F 5’-TCAACCAACCACAAAGACATTGGCAC-3’;

UniCO1-R 5’-TAGACTTCTGGGTGGCCAAAGAATCA-3’;

UniD-loop-F1 5’-CTACCTCCA ACTCCCAAAGC-3’;

UniD-loop-R 5’-CCTGAAGTAGGACCAGATC-3’.

The temperature protocol used for the CO1 PCR replication is 1 cycle of 2 minutes at 95°C; 35 cycles of 30 seconds at 94°C, 30 seconds at 54°C, and 1 minute at 72°C; and 1 cycle of 10 minutes at 72°C (Pegg et al 2006). The temperate protocol used for D-Loop is known as “touchdown” PCR and was 1 cycle of 2 minutes at 94°C; 5 cycles of 30 seconds at 94°C, 30 seconds at 45°C, and 2 minute at 72°C; 5 cycles of 30 seconds at 94°C, 30 seconds at 43°C, and

2 minutes at 72°C; 25 cycles of 30 seconds at 94°C, 30 seconds at 41°C, and 2 minutes at 72°C; and 1 cycle of 10 min at 72°C (Pegg et al 2006). For each reaction, 25µL reactions were run using 12.5µL Master Mix from New England Biolabs (neb.com), 9.5µL highly purified HPLC grade water, .05µL of each of the primers (forward and reverse per gene), and 2µL of the DNA sample. A small amount of PCR optimization was done in which the DNA sample material was increased to 6µL and the amount of water was reduced accordingly to yield a 25µL reaction.

After the PCR was complete, the samples were run on electrophoresis gels to check success of the PCR reaction in replicating a large portion of the target sequence. A 1%TBE gel was used with a 1X SybrSafe DNA Gel Stain. For checking the last 100 samples, a 1X TAE gel buffer was used instead of TBE. Bio-Rad mini gel rigs were used for fast and efficient low volume runs. For each sample, 5µL of DNA was combined with 3µL 6X Blue Loading Dye. After loading the gels with 100 base pair ladder and the samples plus the negative control, the gel rig was run for about 30 min at 80V. After the run was completed, the samples were photographed under UV light with an AlphaImager Mini camera. This method of visualization is much safer than the older technique of using toxic ethidium bromide.

After positive gel results of distinct, clear bands were attained for our samples, the Qiagen QIAquick PCR Purification kit was used to clean up the products in preparation for sequencing (qiagen.com). This method removes residual primers in another fast and efficient kit utilizing spin-columns, buffers, and collection tubes for purification of PCR products. These purified products were then sent to the Clemson University Genome Institute (CUGI) for sequencing. To build the primary original data set, an Applied Biosystems Capillary Array Automated DNA sequencing platform operated by CUGI was used. A well-established, active,

collaborative relationship between the Shedlock Lab and CUGI allowed for discounted priority access to these services, along with full tech support.

To analyze the sequenced DNA, a program called Finch TV was used to assemble, edit, and align the primary data matrices (http://www.geospiza.com). Using this program, the polymorphism and molecular evolutionary patterns observed by multiple CO1 and D-Loop sequence alignments for multiple species were statistically profiled. The NCBI toolbox, especially BLAST-based strategies, were used to compare original DNA sequence data to the global content of publicly available reference data online in GenBank (ncbi.nlm.nih.gov).

Results

The first phase of this research was to create an archive of purified genomic DNA from 300 pelagic marine fish larvae. Of the 300 samples processed, over 100 yielded clean target sequences at the CO1 or D-Loop loci. Our analysis yielded 88 successful CO1 sequences and 19 successful D-Loop sequences. Through analysis of the original sequences and use of BLAST to compare them to the global database of genomic information, a taxonomic diversity of more than 25 distinct families of fishes was found (see Figure 1).

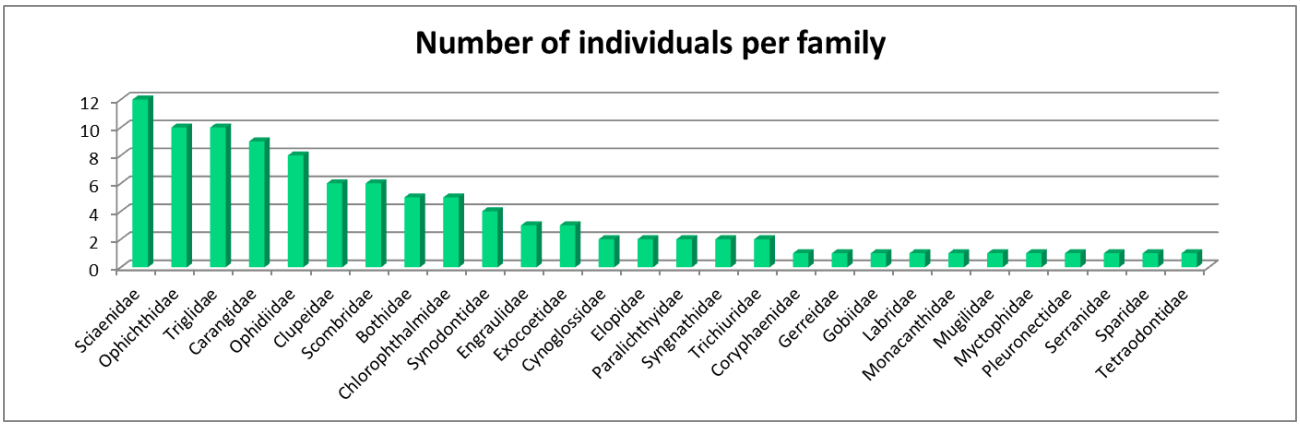


Figure 1. Families of fishes represented in sampling.

Within this broad scope of families, over 35 genera were represented. Top BLAST results for these genera can be found in Appendix 2. The query coverage is a number that tells what percentage of the entered sequence mirrors the DNA bases of the closest match from the GenBank database. For 94% of the CO1 sequences this percentage was 95% or above. Only 49% of the D-Loop sequences had query coverage of 95% or above. The E-value reflects the possibility of the match happening by chance. For 82% of the CO1 sequences, this number was zero. None of the D-Loop sequences had E-values of zero, but all of them were all still exceedingly low values much less than standard cut-off values of ~ E^{-10} to E^{-20}.

From the 20 oceanic sampling stations represented in the samples from the *Transects* program (see Appendix 1), our analysis generated successful target DNA sequence results from 11 stations. For ease of summary, these stations were divided into shallow, medium, and deep ocean oceanographic regions. The shallow stations were at depths of 0-30 meters, the medium stations were at depths of 30-100 meters, and the deep stations were at depths of 100-400 meters (see Figure 2).

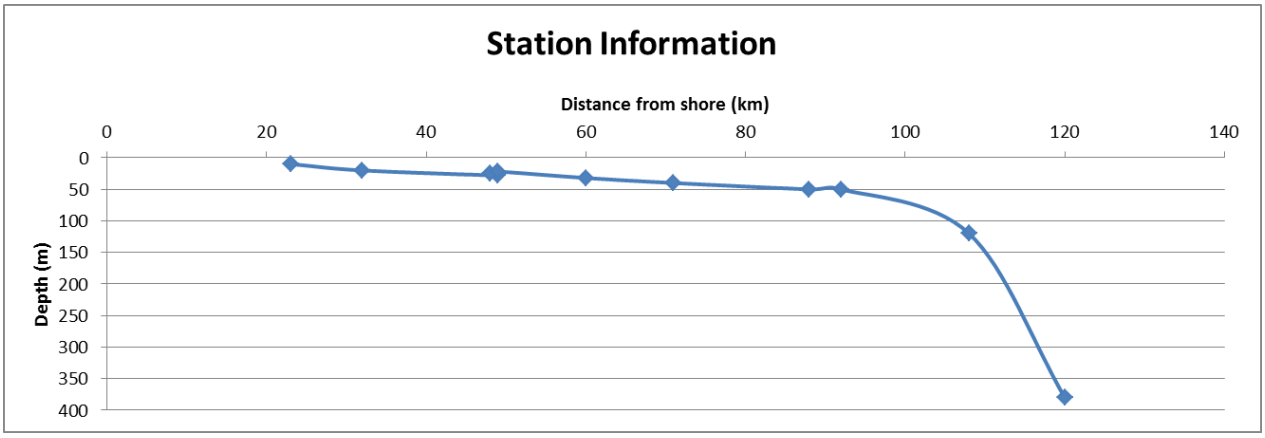


Figure 2. Sampling stations according to depth and distance from shore.

Within the shallow stations, there were 16 families represented among the 31 individuals present (Figure 3). The most common family in the shallow depth stations was Triglidae with 8 individuals. Ophichthidae and Ophidiidae were next in frequency with 4 individuals each. Within

the medium depth stations, there were 14 families represented among the 34 individuals present (Figure 4). The most common family in the medium stations was Sciaenidae with 9 individuals, and Ophichthidae was next most frequent with 4 individuals. Bothidae, Carangidae, and Synodontidae were next most frequent with 3 individuals each. Within the deep stations, there were 15 families represented among the 37 individuals present (Figure 5). The three most frequent families represented in the deep stations were Carangidae, Clupeidae, and Scombridae with 6 individuals each.

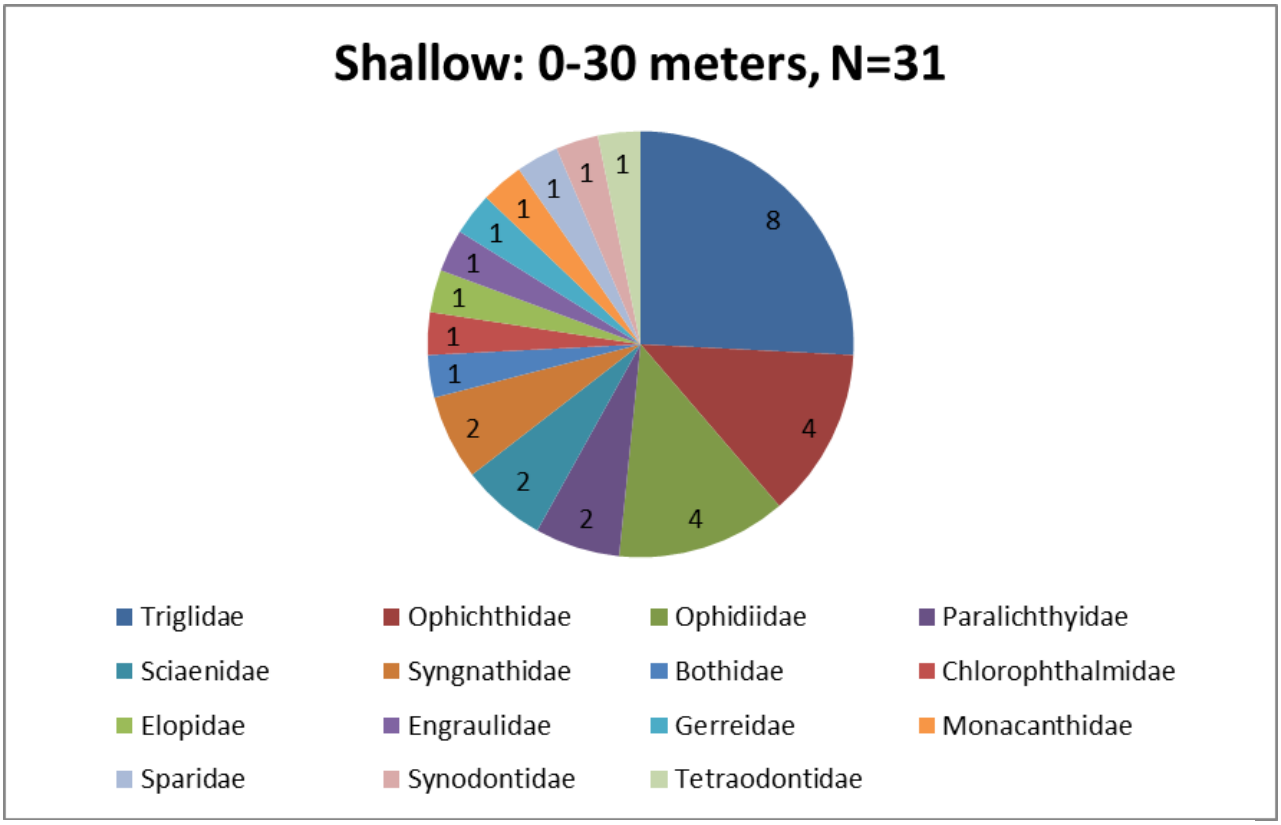


Figure 3. Families represented in shallow sampling stations.

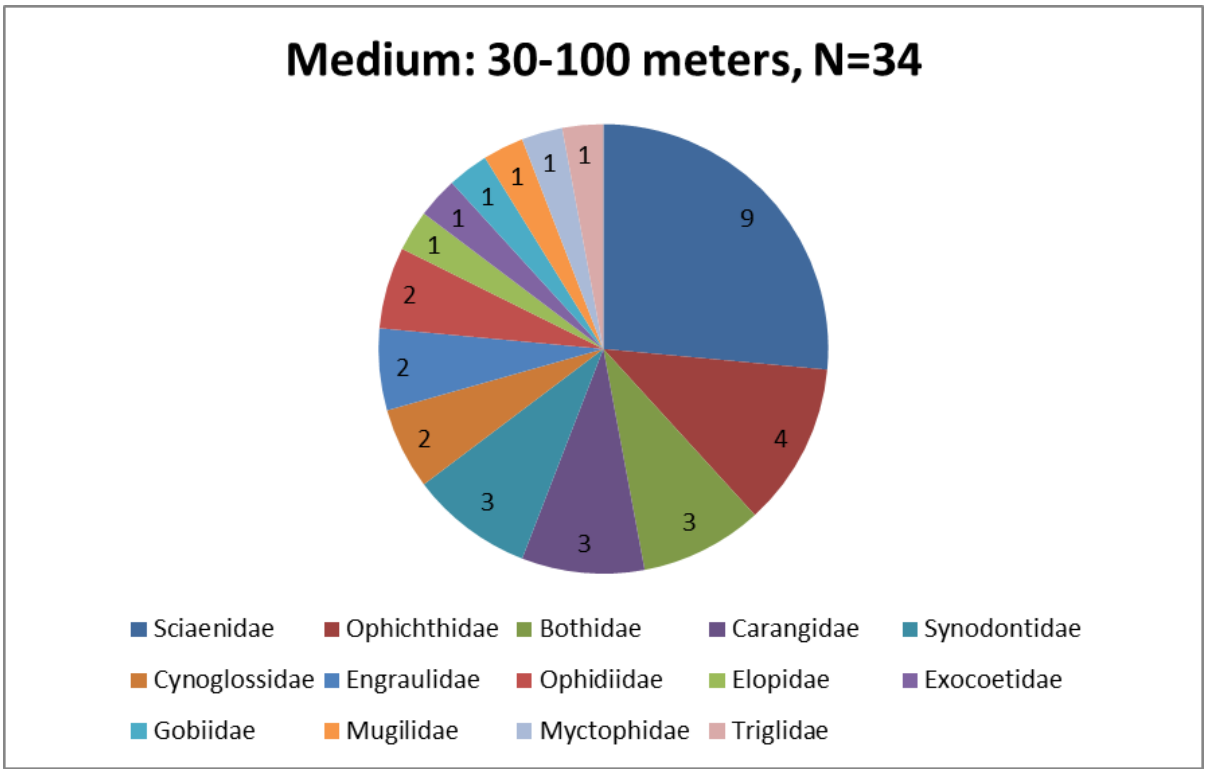


Figure 4. Families represented in the medium depth sampling stations.

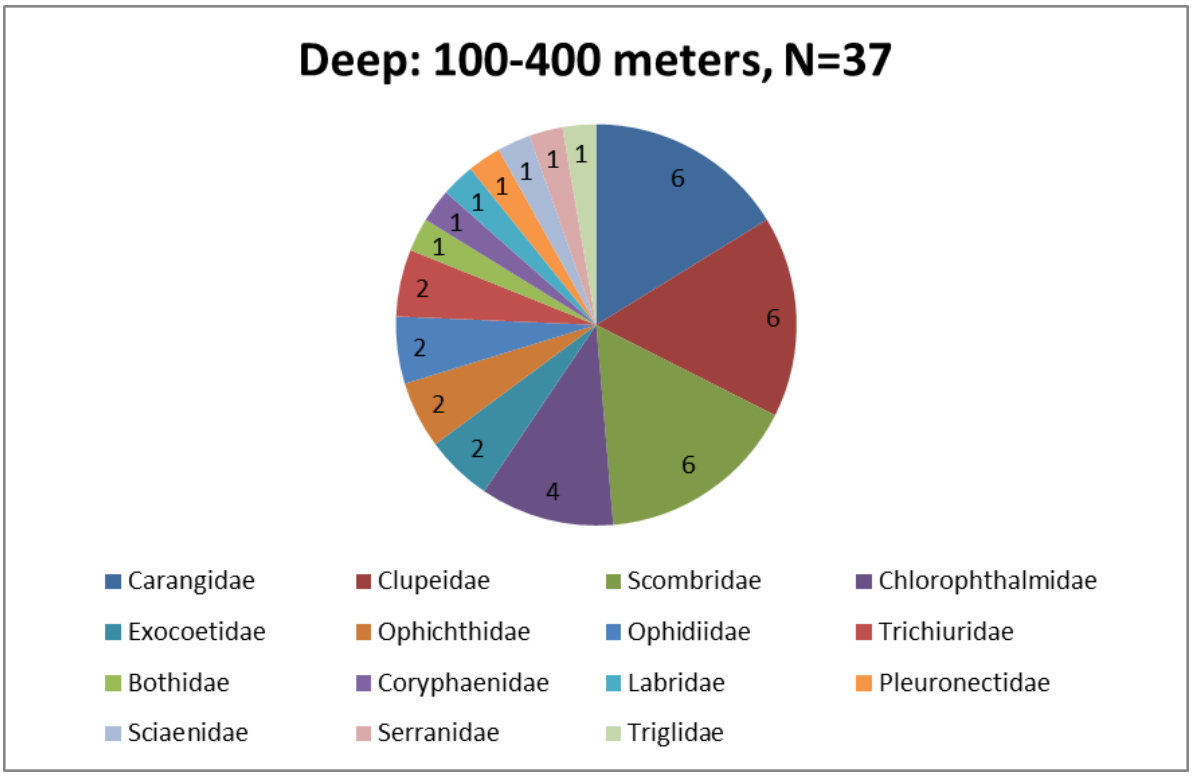


Figure 5. Families represented in the deep sampling stations.

Samples which yielded BLAST results with high query coverage and low E-value were assumed to be correct as long as the matched species is endemic to the continental shelf area off the coast of South Carolina. Some of the samples yielded genus and species results that are not found in this area, so those were assumed to be inaccurate but of a similar family, genus, or species not yet accessioned into the database. Less than 3% of the CO1 sequences returned a taxonomic designation that cannot be found in the sampling area. Within the D-Loop sequences there was a much higher degree of inaccurate or unreliable taxonomic results. Inaccuracies were present in 21% of the family results, in 42% of the genus results, and in 58% of the species results.

In a previous project done by Tarpey (2010), some of the samples were analyzed based on morphology and many were labeled according to possible families. In the present study, preliminary identifications were also estimated visually based on obvious body shape or physical characteristics. Of the total number of morphological estimates made before genetic analysis, less than 10% were correct.

Discussion

There was a wide range of samples present, with their identity mostly unknown due to ambiguous larval morphology. To the best of our ability, a wide taxonomic range of samples was chosen for DNA extraction and sequencing. There was variable success in PCR and usable sequences among the different samples. There were 300 total larvae from which genomic DNA was extracted. During the process of performing PCR for all the samples, including a brief period of PCR optimization, 498 total reactions were run. Of these reactions, 370 were for CO1 and 128 were for D-Loop. Despite initial very high rates of successful PCR reactions, overall

PCR success rate was about 25% (see Table 1). Success was determined if clear bands were visible in the electrophoresis gels. Of these successful PCR reactions that were then purified and sent to CUGI for sequencing, about 90% yielded usable sequences. Failure was due to noisy sequences that did not yield a logical result in BLAST or simply returned a blank sequence from CUGI. This project involved a great deal of manual bench work and optimization to obtain the final archive of frozen purified genomic DNA and the optimized primary data set of larval fish sequences. This physical library of unique voucher specimen DNA and associated digital sequences for two mitochondrial genes are a valuable scientific resource that is transferable to other independent studies by a diversity of investigators.

Total # DNA Archived	300	
PCRs run		
CO1	370	
Dloop	128	
Total	498	
Successful PCR (bands)		**Percent Success compared to total PCRs run**
CO1	89	24.1
Dloop	30	23.4
Total	119	23.9
Successful sequences		**Percent success compared to successful PCR**
CO1	88	98.9
Dloop	19	63.3
Total	107	89.9

Table 1. PCR Success

When analyzing the BLAST results, query coverage and E-value are two important values to consider. High query coverage means more of the inputted query sequence aligns successfully with the bases in the best reference sequence match. Low E-values (with the largest negative exponents) are more significant results because they have a lower possibility of a match occurring by chance. The query coverage of most of the samples in this study was very strong.

Almost all of the CO1 matches had query coverage values of 95% and above. Almost half of the D-Loop matches had query coverage values of 95% and above. A very high percentage of CO1 sequences yielded BLAST results with extremely low E-values; in fact, most of them were zero. Among the D-Loop sequences, the E-values were not quite as strong but they were still much less than 1. The highest E-value among the D-Loop sequences was 3.00×10^{-37} which means there is an exceedingly small probability that even those imperfect matches could have happened by chance. Overall these high query coverage and low E-values strongly suggest that our taxonomic designations are reliable.

Over 25 families were represented in these sequences. This is a very good result because it covers a wide taxonomic range of shallow, benthic fishes as well as open water, pelagic fishes. The top three families in the shallow stations were Triglidae (sea robins), Ophichthidae (snake eels), and Ophidiidae (cusk-eels). These results are consistent with expectations based on ecology and life history in that these 3 families are known to live in shallow water, especially the eels which typically live in association with the seafloor. Other near-shore families such as puffers and filefish were found only at shallow stations. The families found only in the shallow stations were Gerreidae, Monacanthidae, Paralicthyidae, Sparidae, Syngnathidae, and Tetraodontidae.

At the medium depth stations Sciaenidae (drums, croakers) and Ophichthidae were the most common families. Once again, these results are consistent with life history characteristics as these fishes typically live in shallow to mid-depth waters, not out in the deep ocean. A surprising result from the medium stations was the presence of one member of the family Myctophidae (lanternfish; CO1, 100% query coverage, E-value = 0), which is typically a

mesopelagic fish that does not live near the surface. The families found only in the medium depth stations were Cynoglossidae, Gobiidae, Mugilidae, and Myctophidae.

In the deep stations, the top families were Carangidae (jacks), Clupeidae (sardines, herrings), and Scombridae (mackerels, tunas). These families were not surprising because they are known to be deeper water fishes. Several pelagic families such as mahi-mahi and tunas were found only from deep stations. The families found only at deep stations were Clupiedae, Coryphaenidae, Labridae, Pleuronectidae, Scombridae, Serranidae, and Trichiuridae.

Lack of heavily skewed taxonomy by depth suggests the community is well mixed across variable oceanographic conditions spanning the continental shelf. Some supporting evidence for this statement is that families such as snake eels, cusk-eels, sea robins, drums, and lefteye flounders were found in shallow, medium, and deep stations. The percentage of families shared among all depths was 17.9%, indicating a mixed community. There was a great deal of overlap of families among all depths, but the next largest percentage of total families was shared between shallow and medium depth regions at 10.7%. The percentage of families shared between medium and deep stations was 7.1%, and the percentage shared between shallow and deep stations was only 3.6%.Figure 6 summarizes results for genotyped families integrated with oceanographic conditions in a Venn diagram.

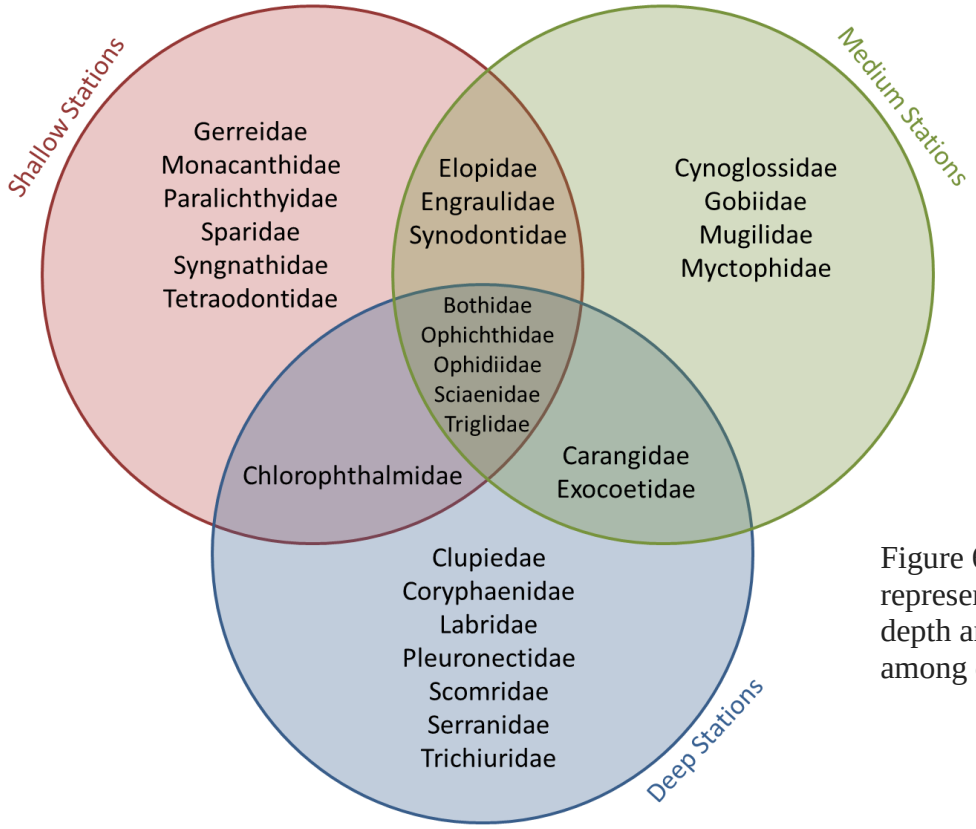


Figure 6a. Families represented at each depth and shared among depths.

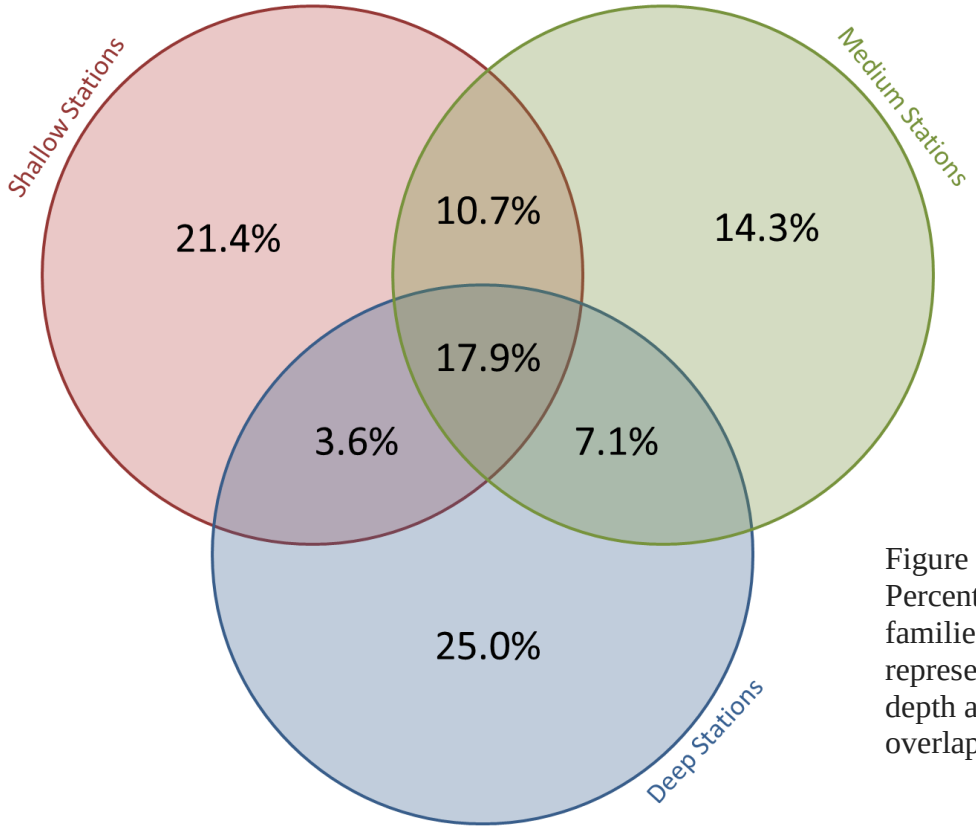


Figure 6b. Percentages of families represented at each depth and at overlapping depths.

This synoptic view of distinct vs. overlapping patterns can be explained in part by the sharp depth gradient in continental slope waters near the outer limits of our sampling transect area, as reflected by the graph in Figure 2 and the isolines for depth mapped off the shores of South Carolina in Appendix 1. It is expected that further analysis of more of the samples collected at all stations will lead to more total family diversity, but that the patterns of dominant families identified by the present study between shallow-medium vs. deep water will be proportionally reinforced. In fact, 25% of the families were found only at the deep stations, which is the highest percentage of unique families among the varying depth summaries. This trend is likely due to selection for life history traits associated with regional oceanographic depth profiles and benthic vs. mid-water eco-morphology. It is also expected that further forensic DNA typing of more individuals will show that some non-dominant families present only at certain stations in the investigation may occur over a more widespread area. Continued analysis should lead to better resolution of distinct trends and patterns of larval recruitment and fish community diversity in the waters surrounding South Carolina.

Determining dispersal is valuable in predicting or detecting environmental changes due to anthropogenic pressures such as pollution, current dynamics, ocean acidification, and climate change impacts on average seasonal water temperatures. Our work is relevant and important to other professional researchers such as ichthyologists and fisheries biologists. These successful genotyping results will continue to be integrated with physical oceanography and biogeography to help bridge a large gap in our ability to model juvenile recruitment and community dynamics of ecologically important pelagic fish populations. Our analysis revealed common types of larvae present in the Western Atlantic at multiple depths, which is information that can be significant to many different fields such as fisheries biology, marine resource management, and climate change

research. This work also greatly strengthens the scientific value of the *Transects* program at the College of Charleston through continued analysis of samples and locality data.

Not all samples yielded definitive results to genus and species level when entered into BLAST. This result likely means that the specific species queried is not yet in the database. Among the samples, there was a much higher degree of accurate, reliable taxonomic designations for CO1 than for D-Loop. This result was anticipated and is likely due to the bias in using CO1 much more often than D-Loop for forensic identification. D-Loop is a well-established polymorphic region that has been valuable for population genetic studies of commercially important species but is still a fairly new genetic marker for forensic genotyping of a broad taxonomic diversity of fishes. As such, D-Loop has a much smaller fraction of accessioned reference sequences in the global database as compared to CO1. Also, D-Loop is a non-coding region that can be extremely variable among different species, so there may be too much variation across divergence levels between some queried species and their closest relatives in the reference database to determine precise species designation, even if there were more original query information available.

This bias in the D-Loop database made it more difficult to definitively identify our samples, but it also provides the opportunity to strengthen the reference database through accessioning our original CO1 and especially D-Loop sequences into GenBank. This will be particularly valuable as more independent genomic loci such as CO1 and nuclear genes can be associated with D-Loop results for the same samples in the voucher-based archive of purified genomic DNAs. The present research has both relied upon and can strengthen the value of the public database available to the global scientific community.

Conclusions

During this research, the proof-of-concept that forensic identification of larval fish is much more reliable than morphological analysis was demonstrated. Moreover, the use of two loci, CO1 and D-Loop, provides clear advantages over single-locus approaches to forensic genotyping, especially in strengthening the diagnostic power of the public reference database.

The results indicate that depth of habitat and associated life history can in part explain preliminary larval fish taxonomic distributions across the continental shelf of South Carolina. With further genotyping of additional individuals at multiple genomic loci, it is expected that overall diversity would increase and that resolution of trends of larval dispersion would become more evident and more finely resolved by region and oceanographic condition. It is anticipated that pelagic open water species such as top predators would continue to be identified primarily from the deep stations with more mixed taxonomic distributions shared across shallow to medium depth coastal habitats. Original DNA sequences generated from this study will be submitted as a manuscript for publication in a peer-reviewed journal and accessioned into GenBank for public scientific use.

Acknowledgements

We thank Gorka Sancho from Grice Marine Lab for generously providing samples from the *Transects* program and Dr. Chris Saski from Clemson University for DNA sequencing support. This work was supported in part by a College of Charleston MAYS grant provided to A. Cole by the Office of Undergraduate Research and by an Innovative Teaching and Learning Grant awarded to A. Shedlock by the Office of Academic Affairs. A. Cole would like to emphasize that this project would not have been possible without the support and guidance of her undergraduate research mentor Dr. Shedlock.

Literature Cited

Bignami, S., Sponaugle, S., and Cowen, R.K. 2013. Response to ocean acidification in larvae of a large tropical fish, *Rachycentron canadum*. Global Chance Biology 19: 996-1006.

BLAST. http://blast.ncbi.nlm.nih.gov/Blast.cgi

FinchTV. http://www.geospiza.com/Products/finchtv.shtml

GenBank. http://www.ncbi.nlm.nih.gov/genbank/

Gomes, C., Dales, R. G. B, & Oxenford, H. A. (1997). The use of DNA markers in stock discrimination of commercially important fish species and their application to management of the Four-wing Flying Fish, *Hyrundichthys affinis*, in the Caribbean.

Hebert, P.D., Ratnasingham, S., & deWaard, J.R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings: Biological Sciences* 270: 96-99.

IDT. http://www.idtdna.com/site

Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Paabo, S., Villablanca, F.X., & Wilson, A.C. (1989). Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci.* 86: 6196-6200.

McKinley, A.C., Miskiewicz, A., Taylor, M., and Johnston, E. 2011. Strong links between metal contamination, habitat modification and estuarine larval fish distributions. Environmental Pollution 159: 1499-1509.

New England Biolabs. www.neb.com

Oceanica, College of Charleston. "Map of Charleston Transects." *Project Oceanica*. Web. 27 Oct. 2012. <http://oceanica.cofc.edu/transects/home.htm>.

Palumbi, S., Martin, A., Romano, S., McMillan, W.O., Stice, L., Grabowski, G. (2002). The simple fool's guide to PCR Version 2.0. Compiled by Department of Zoology and Kewalo Marine Laboratory, University of Hawaii.

Pegg, G.G., Sinclaire, B., Briskey, L., & Aspden, W.J. (2006). MtDNA barcode identification of fish larvae in the southern Great Barrier Reef, Australia. *Scientia Marina* 70S2: 7-12.

Qiagen. www.qiagen.com

Richardson, D.E., & Cowen, R.K. (2004). New leptocephalus types collected around the island of Barbados (West Indies). *Copeia* 2004(4): 888-895.

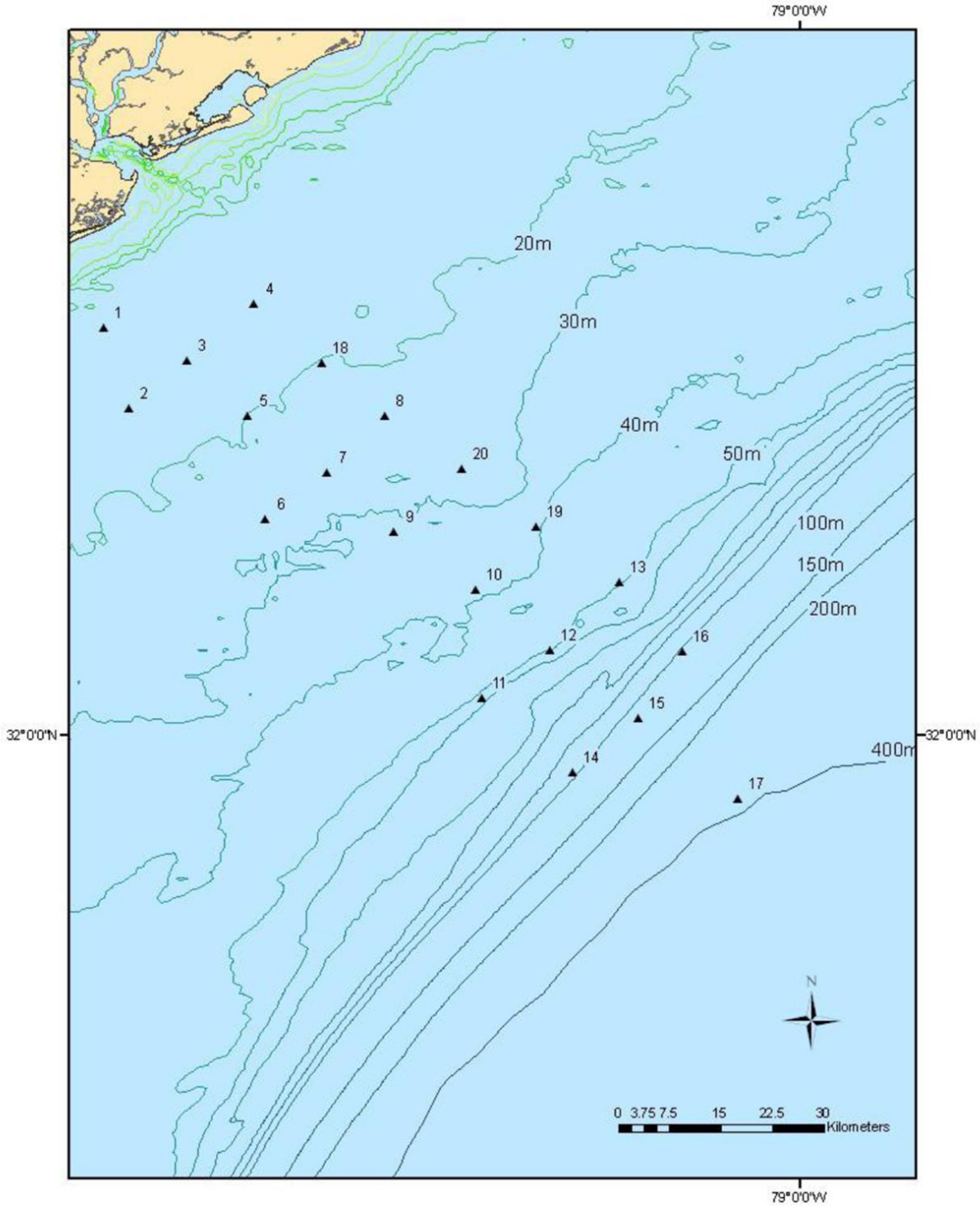
Savolainen, V., Cowan, R.S., Vogler, A.P., Roderick, G.K., Lane, R. (2005). Towards writing

the encyclopedia of life: an introduction to DNA barcoding. *Philosophical Transactions: Biological Sciences* 360: 1805-1811.

Tarpey, C. (2010). Morphological and genetic identification of Western Atlantic bight larval fishes.

Vences, M., Thomas, M., Bonett., Vieites, D.R. (2005). Deciphering amphibian diversity through DNA barcoding: chances and challenges. *Philosophical Transactions: Biological Sciences* 360: 1859-1868.

Ward, R. D., Zemlak, T.S., Innes, B.H., Last, P.R., & Hebert, P.D.N. (2005). DNA barcoding fish species. *The Philosophical Transactions of the Royal Society B* 360, 1847-1857.



Appendix 1. Map of sampling stations off the coast of South Carolina.

CO1				
Sample Number	**Family**	**Genus**	**Query Coverage (%)**	**E-value**
233	Bothidae	Bothus	99	3.00E-146
266	Carangidae	Decapterus	100	0
267	Carangidae	Selar	97	0
104	Carangidae	Seriola	100	0
256	Carangidae	Trachinotus	98	0
100	Chlorophthalmidae	Chlorophthalmus	100	0
103	Clupeidae	Sardinella	100	0
105	Coryphaenidae	Coryphaena	99	0
172	Cynoglossidae	Symphurus	97	0
21	Elopidae	Elops	98	2.00E-84
140	Engraulidae	Engraulis	100	0
40	Exocoetidae	Cheilopogon	84	0
280	Exocoetidae	Hirundichthys	98	0
73	Gerreidae	Eucinostomus	99	7.00E-112
89	Gobiidae	Microgobius	95	0
232	Labridae	Xyrichthys	100	2.00E-145
294	Monacanthidae	Stephanolepis	100	8.00E-118
63	Mugilidae	Mugil	100	3.00E-67
33	Myctophidae	Hygophum	100	0
13	Ophichthidae	Myrophis	96	0
48	Ophichthidae	Ophichthus	100	0
87	Ophidiidae	Ophidion	97	0
101	Paralichthyidae	Etropus	99	0
72	Sciaenidae	Leiostomus	99	0
178	Sciaenidae	Micropogonias	100	0
241	Scombridae	Auxis	100	0
80	Scombridae	Katsuwonus	97	0
281	Serranidae	Serranus	98	0
262	Syngnathidae	Hippocampus	99	0
128	Synodontidae	Synodus	100	0
52	Tetraodontidae	Sphoeroides	99	0
96	Trichiuridae	Trichiurus	100	0
71	Triglidae	Prionotus	100	0
Dloop				
Sample Number	**Family**	**Genus**	**Query Coverage (%)**	**E-value**
104	Carangidae	Seriola	99	7.00E-95
246	Carangidae	Trachurus	100	1.00E-99
103	Clupeidae	Sardinella	95	1.00E-138
48	Ophichthidae	Ophisurus	99	5.00E-85
197	Pleuronectidae	Verasper	47	2.00E-38
127	Sciaenidae	Cynoscion	99	3.00E-55
241	Scombridae	Auxis	99	1.00E-161
80	Scombridae	Katsuwonus	96	3.00E-170
52	Sparidae	Diplodus	75	3.00E-37
173	Synodontidae	Synodus	96	1.00E-53

Appendix 2. Summary of BLAST results by genera.