DNA Barcoding: A Genetic and Morphological Analysis of Three Closely Related Notropis Shiners.

Amanda Crespo, Michael Brett, Jo-Jo Kelly and Chris Lorentz, Ph.D.

Thomas More University, 333 Thomas More Parkway, Crestview Hills, KY, 41017

Thomas More University, Thomas More Biology Field Station and Center for Ohio River Research and Education, 8309 Mary Ingles Highway, California, KY, 41007

Universidad del Sagrado Corazón, Calle San Antonio Rosales, Parada 26.5, San Juan, 00914, Puerto Rico

Abstract

Morphological features of fish have been one of the main references for fish identification (ID), but it is unreliable since many share physical features that lead to misidentification. Instead, a technique called DNA Barcoding can be used to identify species according to their gene-specific DNA sequence. DNA barcoding requires the cytochrome c oxidase (CO1) gene, which is a mitochondrial gene most abundantly found in smooth and skeletal muscles (Bhattacharya et al., 2015). The DNA sequence of the CO1 gene is unique to each species, which makes it an ideal marker for species identification. In this study, ten samples of three shiner fish species were collected from the Ohio River: *Notropis atherinoides* (Emerald Shiners), *Notropis blennius* (River Shiners), and *Notropis volucellus (*Mimic Shiners). These were morphologically identified, then their pectoral fin tissue was extracted and used for the DNA sequencing. This study aims to differentiate the identification of three minnow species using morphological and genetic variations. It was predicted that the study's morphological identification would match the DNA barcoding identification.

Introduction

Morphological features of fish have been one of the main references for fish identification (ID), but it has been shown to be unreliable since many share physical features that lead to misidentification. Instead, a technique called DNA Barcoding can be used to identify species according to their gene-specific DNA sequence. DNA barcoding requires the cytochrome c oxidase (CO1) gene, which is unique to each species, making it an ideal marker for species identification. FISH-BOL is the campaign to create a global reference library, the Barcode of Life Database System (BOLD), with a plethora of sequences for various fish species (Hanner et al 2011). Specimens gathered in the field can be sequenced and submitted to BOLD for species identification. This study's contribution to BOLD will allow a greater understanding of fish genetic dynamics in the Ohio River Basin and around the globe (Becker et al 2011). The specimens used for this study were ten samples of three shiner fish species from the Ohio River: *Notropis atherinoides* (Emerald Shiners), *Notropis blennius* (River Shiners), and *Notropis volucellus* (Mimic Shiners). These were morphologically identified, then their pectoral fin tissue was extracted and used for the DNA sequencing. The aim of this study is to differentiate the identification of three minnow species using morphological and genetic variations. It was predicted that the study's morphological identification would match the DNA barcoding identification.

Methodology

The fish were collected through electrofishing from the Ohio River and 12-mile Creek before they were examined for the following characteristics: size, color, the number of rays present on the anal fin, and the number of lateral line scales. The goal was to use morphological analysis to identify the species of shiner before genetic sequencing began.

The genetic sequencing began with DNA extraction where 1 mm² of pectoral fin tissue was put in a 1.5 mL tube with 600 µL of digestion buffer and 20 µL of proteinase K. It was incubated overnight, followed by the addition of 500 µL of phenol chloroform to each tube and put into the centrifuge at 13,000 rpm for 10 minutes. The top layer was removed and put into clean 1.5 mL tubes followed by another 400 µL of phenol chloroform and 10 minutes in the centrifuge. Afterwards 1000 µL of Ethanol and 100 µL of sodium acetate was added to each tube and placed in the freezer overnight. Lastly, the samples were run through the centrifuge for 30 minutes before 1 mL of 70% Ethanol was inserted into the solution with another 10 minutes in the centrifuge and the addition of 50 µL of molecular grade water.

The Polymerase Chain Reaction (PCR) targeted the CO1 gene using the CO1-3 primer set using a stock of the primer cocktail. The recipe for the stock solution is as follows: 30 µL nanopure PCR water, 5 µL of VF2_t1, FishF2_t1. FishR2_t1, FR1d_t1, and 100 µL of Taq Master Mix. The samples were prepared with 180 µL of TE, 7.5 µL of Master Mix, and 2.5 µL of DNA before being put into the PCR machine and run on the program ST-50 for 2.5 hours as shown in Figure 1. To continue, Gel electrophoresis was used to confirm that the PCR worked and that the gene was indeed isolated as shown in Figure 2. The first clean-up process began by mixing 2 µL of the EXOSAP-IT reagent with 5 µL of the PCR product, which was them incubated 35 for 15 minutes; later, the samples were incubated at 80 for 15 minutes to inactivate the EXOSAP-IT reagent. Following the last incubation period was a second round of PCR where 1.5 µL of clean DNA was mixed with 8.5 µL of the Master Mix, which was created for each forward primer (M13F) and each reverse primer (M13R), using the following Big Dye Master Mix Recipe: 115 µL of nanopure PCR water, 10 µL of primer, 40 µL of buffer, and 5 µL of Big Dye. The second cleaning began with the addition of 15 µL of SAM solution followed by 3 µL of beads and the samples were vortexed before the removal of 20 µL of DNA solution. The sequencing plate was then put into the ABI 3500 Genetic Analyzer for sequencing.

The sequencing reactions were run on the aforementioned genetic analyzer using the M13F and M13R primers, Figure 3. These sequences were then aligned using the software program known as Geneious and were then combined into a single harmonious sequence. The primer sequences were trimmed from each sequence to generate the final barcoding sequence that was to be analyzed. This is the step where the sequences produced by the sequencer can be compared to those on public databases, specifically the Barcode of Life Data Systems (BOLD) in the case of this experiment, for identification. Once each sample was identified based off the CO1 gene sequence, these results were compared to what the species was thought to be based off the morphology.

Methodology

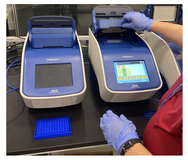

Figure 1.
PCR Machine

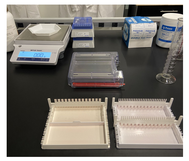

Figure 2.
Gel Electrophoresis

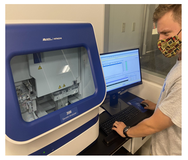

Figure 3.
ABI 3500 Genetic Analyser

Results

The following are images that represent one of the three species that were morphologically identified as N. atherinoides, N. blennius, and N. volucellus, and later used for the DNA Barcoding.

Figure 4.
Notropis atherinoides

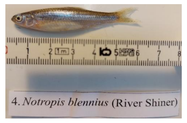

Figure 5.
Notropis blennius

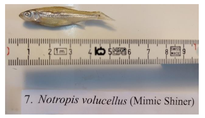

Figure 6.
Notropis volucellus

The DNA sequences matched our morphological identification for 16 of our specimens, while the other 11 samples did not. The *Notropis atherinoides* Sample Group (S.G.) were all positively identified as *N. atherinoides* with a 99.92% on average (X) match for all specimens as shown in Table 01.The second closest match for *N. atherinoides* S.G. was *N. stilbius* (Silverstripe Shiner) 97.64% X, and the third match was *N. amoenus* (Comely Shiner) with 97.08%.

For the *Notropis blennius* S.G., 6 out of 7 samples were identified as *N. blennius* at 99.98% X. Sample 1 of the *N. blennius* S.G. matched as *N. girardi* at 99.85% as shown in Table 02. The second closest match was N. pottery (Chub Shiner) with 98.14% X, and the third match was N. spectrunculus (Mirror Shiner) with a 94. 65 % X match.

The *Notropis volucellus* S.G. all came back as *N. girardi* in the BOLD database with 99.71% . The second closest match for the *N. volucellus* S.G. was *N. buchanani* (Ghost Shiner) 97.98% X. The third closest match was *N. spectrunculus* (Mirror Shiner) with 97.81% X, and two were *N. volucellus 97.78%* X as shown in Table 03 below.

Table 01: Match % of *N. atherinoides* S.G. to fish in the BOLD System

BOLD Percentage match:
Notropis atherinoides (Emerald Shiner) Sample Group

	Primary Match	%	Secondary Match	%	Tertiary Match	%
Mean		**99.924**		**97.639**		**97.077**
Sample 1	N. *atherin.*	100	N. *stilbius*	97.59	N. amoe.	97.09
Sample 2	N. *atherin.*	100	N. *stilbius*	97.93	N. amoe.	97.09
Sample 3	N. *atherin.*	99.54	N. *stilbius*	97.24	N. amoe.	96.77
Sample 4	N. *atherin.*	100	N. *stilbius*	97.76	N. amoe.	97.26
Sample 5	N. *atherin.*	100	N. *stilbius*	97.76	N. amoe.	97.26
Sample 6	N. *atherin.*	99.85	N. *stilbius*	97.41	N. amoe.	96.93
Sample 7	N. *atherin.*	100	N. *stilbius*	97.59	N. amoe.	97.09
Sample 8	N. *atherin.*	99.85	N. *stilbius*	97.76	N. amoe.	96.93
Sample 9	N. *atherin.*	100	N. *stilbius*	97.76	N. amoe.	97.26
Sample 10	N. *atherin.*	100	N. *stilbius*	97.59	N. amoe.	97.09

Table 02: Match % of *N. blennius* S.G. to fish in the BOLD System

BOLD Percentage match:
Notropis blennius (River Shiners) Sample Group

	Primary Match	%	Secondary Match	%	Tertiary Match	%
Mean		**99.96**		**98.14**		**94.65**
Sample 1	N. *girardi*	99.85	N. *Buchan.*	98.03	N. *spectrun.*	97.85
Sample 2	N. *blennius*	100	N. *potteri*	98.16	N. *edward.*	94.04
Sample 3	N. *blennius*	100	N. *potteri*	98.16	N. *edward.*	94.04
Sample 4	N. *blennius*	100	N. *potteri*	98.16	N. *edward.*	94.2
Sample 5	N. *blennius*	99.85	N. *potteri*	98.15	N. *edward.*	94.19
Sample 6	N. *blennius*	100	N. *potteri*	98.16	N. *edward.*	94.04
Sample 7	N. *blennius*	100	N. *potteri*	98.16	N. *edward.*	94.2

Results

Table 03: Match % of *N. volucellus* S.G. to fish in the BOLD System

BOLD Percentage match:
Notropis volucellus (Mimic Shiner) Sample Group

	Primary Match	%	Secondary Match	%	Tertiary Match	%
Mean		**99.709**		**97.982**		**97.805**
Sample 1	N. *girardi*	99.85	N. *buchan.*	98.03	N. *spectrun.*	97.85
Sample 2	N. *girardi*	99.54	N. *buchan.*	97.87	N. *spectrun.*	97.85
Sample 3	N. *girardi*	99.85	N. *buchan.*	98.03	N. *spectrun.*	97.85
Sample 4	N. *girardi*	99.85	N. *buchan.*	98.03	N. *spectrun.*	97.85
Sample 5	N. *girardi*	99.69	N. *buchan.*	98.19	N. *volucellus*	97.85
Sample 6	N. *girardi*	99.85	N. *buchan.*	98.03	N. *spectrun.*	97.85
Sample 7	N. *girardi*	99.69	N. *buchan.*	97.87	N. *spectrun.*	97.7
Sample 8	N. *girardi*	99.85	N. *buchan.*	98.03	N. *spectrun.*	97.85
Sample 9	N. *girardi*	99.69	N. *buchan.*	97.87	N. *spectrun.*	97.7
Sample 10	N. *girardi*	99.23	N. *buchan.*	97.87	N. *volucellus*	97.7

Conclusion

Morphological ID for fish is partially convenient, but it is an error-prone strategy in comparison to genetic ID. Results show that Sample 1 of the *N. blennius* S.G. matched with an *N. girardi,* and because all the N. volucellus matched with *N. girardi*, it could be implied that Sample 1 of *N. blennius* S.G. was misidentified before the lab work took place. In addition, the *N. volucellus* S.G. matched with the Arkansas River Shiner, which is conflicting because geographically they are found in the Canadian River in New Mexico, Oklahoma, Texas, and Arizona, which is distant from the Ohio River (Shiner, River, Fenner, & Usfws, n.d.). The BOLD database currently has 26 Mimic Shiners in the public record (July/2021), therefore, the samples of this study should have matched with their data. Considering all of the Mimic shiner specimens match as Arkansas River Shiners, this indicates genetic relation. Sample 5 of the *N. volucellus S.G.* tertiary match was related to N. volucellus with a 97.85%, therefore, it is still a possibility that it is a Mimic Shiner. Overall, high genetic similarities (99.00% to 97.00% X̄) between Primary, Secondary, and Tertiary matches could be misidentified if one relies on morphological ID. Morphological identification can be successful between *N. atherinoides* and *N. blennius* species. DNA Barcoding remains a useful tool for accurately identifying species.

References

Bhattacharya M, Sharma AR, Patra BC, Sharma G, Seo E-M, Nam J-S, Chakraborty C, Lee S-S. 2015 Jun 9. DNA barcoding to fishes: current status and future directions. Mitochondrial DNA.:1–9. doi:10.3109/19401736.2015.1046175.

Shiner A, River A, Fenner S, Usfws. 2011. Arkansas River Shiner Oklahoma Ecological Service Field Office. https://www.fws.gov/southwest/es/oklahoma/Documents/TE_Species/Species%20Profiles/AR%20River%20Shiner.pdf

Steinke D, Hanner R. 2011. The FISH-BOL collaborators' protocol. Mitochondrial DNA. 22(sup1):10–14. doi:10.3109/19401736.2010.536538.

Ward RD, Hanner R, Herbert PDN. 2008. Review Paper: The campaign to DNA barcode all fishes, FISH-BOL. Journal of Fish Biology. 74:329–356. doi:doi:10.1111/j.1095-8649.2008.02080.x, available online at http://www.blackwell-synergy.com.

Identification Request | BOLDSYSTEMS. (n.d.). Retrieved July 26, 2021, from www.boldsystems.org website: https://www.boldsystems.org/index.php/IDS_OpenIdEngine

International Barcode of Life Consortium. (2017). DNA barcoding - International Barcode of Life. Retrieved from International Barcode of Life website: https://ibol.org/about/dna-barcoding/

Acknowledgements

We would like to thank Dr. Heather Farrington at the Cincinnati Natural history Museum at Union Station for being our mentor throughout the process of sequencing the DNA samples in the museum's genetic lab. Next, we would also like to thank Dr. Robert Hanner for talking with us about Fish DNA barcoding. Finally, we would like to thank the members of the River Crew for helping us catch the shiners for this experiment and Luis Tron for helping us prepare our samples to go to the museum for analysis.

For further information, please contact us:

Amanda Crespo amandacrespo003@gmail.com, Jo-Jo Kelly gfkell32@thomasmore.edu, and Michael Brett mlbret17@thomasmore.edu.