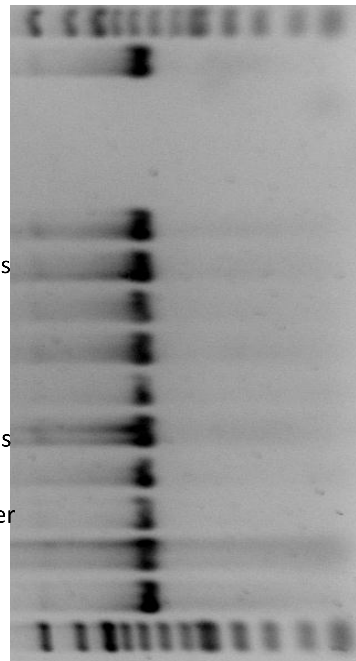
RECLAMATION

Managing Water in the West

Technical Memorandum No. 86-68220-14-14

Identification of Unknown Organisms by DNA Barcoding: A Molecular Method for Species Classification

Research and Development Office
Invasive Mussels
Final Report 2014-01 (0045)



Jacque Keele, Jamie Carmon, Sherri F. Pucherelli, Denise Hosler

Mission Statements

The U.S. Department of the Interior protects America's natural resources and heritage, honors our cultures and tribal communities, and supplies the energy to power our future.

The mission of the Bureau of Reclamation is to manage, develop, and protect water and related resources in an environmentally and economically sound manner in the interest of the American public.

REPORT DOCUMENTATION PAGE		*Form Approved OMB No. 0704-0188*
T1. REPORT DATE September 2014	**T2. REPORT TYPE** Research	**T3. DATES COVERED** 2014 FY
T4. TITLE AND SUBTITLE **Identification of Unknown Organisms by DNA Barcoding: A Molecular Method for Species Classification**		**5a. CONTRACT NUMBER** RY1541ZQ201410045
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER 1541 (S&T)
6. AUTHOR(S) Jacque Keele, jkeele@usbr.gov, 303-445-2187; Jamie Carmon; Sherri F. Pucherelli, spucherelli@usbr.gov, 303-445-2015; Denise Hosler, dhosler@usbr.gov, 303-445-2195		**5d. PROJECT NUMBER** 0045
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER 86-68220
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Bureau of Reclamation, Technical Service Center, Environmental Applications and Research Group		**8. PERFORMING ORGANIZATION REPORT NUMBER** 86-68220-14-14
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Research and Development Office; U.S. Department of the Interior, Bureau of Reclamation, PO Box 25007, Denver CO 80225-0007		**10. SPONSOR/MONITOR'S ACRONYM(S)** R&D: Research and Development Office; BOR/USBR: Bureau of Reclamation; DOI: Department of the Interior
		11. SPONSOR/MONITOR'S REPORT NUMBER(S) 2014-01-0045

12. DISTRIBUTION / AVAILABILITY STATEMENT
Final report can be downloaded from Reclamation's website: https://www.usbr.gov/research/

13. SUPPLEMENTARY NOTES

14. ABSTRACT *(Maximum 200 words)*
DNA barcoding is a molecular method for species identification. The goal of this research was to identify ways that researchers at Reclamation can use this technique to enhance their research. This report includes a literature review of DNA barcoding methods, a summary of some of the barcoding projects performed, and a summary of the lessons learned from the DNA barcoding research done this year. A variety of organisms (birds, fish, mammals, invertebrates) were analyzed this year with varying degrees of success. DNA barcoding can be relatively straightforward (sample collection, isolation of DNA, amplification of the barcoding gene, and finally sequencing and analyzing the results), but there are issues that can occur (poor DNA or the PCR failed to amplify) resulting in the need to troubleshoot and determine what is going wrong. Overall, DNA barcoding supports the work done by field biologists by giving them a way to confirm through molecular biology the identification of their organisms.

15. SUBJECT TERMS Polymerase chain reaction, DNA barcoding, species identification, DNA analysis

16. SECURITY CLASSIFICATION OF: U			**17. LIMITATION OF ABSTRACT** U	**18. NUMBER OF PAGES** 26	**19a. NAME OF RESPONSIBLE PERSON** Jacque Keele
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U			**19b. TELEPHONE NUMBER** 303-445-2187

S Standard Form 298 (Rev. 8/98)
P Prescribed by ANSI Std. 239-18

PEER REVIEW DOCUMENTATION

Project and Document Information

Project Name Identification of Unknown Organisms by DNA Barcoding: A Molecular Method for Species Classification **WOID** X0045

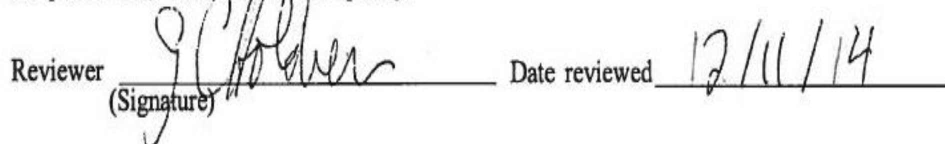
Document Identification of Unknown Organisms by DNA Barcoding: A Molecular Method for Species Classification

Document Author(s) Jacque Keele, Jamie Carmon, Sherri F. Pucherelli, and Denise Hosler
Document date September, 2014

Peer Reviewer: Chris Holdren

Review Certification

Peer Reviewer: I have reviewed the assigned items/sections(s) noted for the above document and believe them to be in accordance with the project requirements, standards of the profession, and Reclamation policy.



Reviewer (Signature) Date reviewed 12/11/14

Acknowledgements

The authors of this research would like to thank the Research and Development Office for supporting and providing funding for this study. Also, thank you to the reviewers of this report and to Kevin Bloom and Kyle Rulli for assisting with this research. Finally, the authors want to thank all of the researchers at Reclamation who provided samples for analysis.

Executive Summary

DNA barcoding is a useful technique for the molecular identification of organisms. This technique uses PCR to amplify a fragment of the cytochrome oxidase I (COI) gene, which is then sequenced and compared to a database of known organisms. For plants, other genes such as rbcL and matK are commonly used for their barcoding identification. The goal of this research was to obtain the necessary polymerase chain reaction (PCR) primers and reagents so that DNA barcoding could be performed on a wide range of taxonomic groups. The PCR primers for invertebrates, fish, mammals, and plants were identified and ordered, and the DNA barcoding methods and protocols for these taxonomic groups were researched in the published literature. Samples from a wide range of organisms were obtained from Reclamation researchers and tested using DNA barcoding methods. Once the PCR reaction was completed, the resulting PCR product was sent for DNA sequencing and analyzed using both the DNA BOLD and NBCI databases. It was possible to analyze DNA samples from fish, birds, mammals, plants, and invertebrates. After samples that had been submitted by Reclamation researchers were analyzed, the results were reported to the researchers. Several lessons about DNA barcoding were learned over the course of this project. This report includes a literature review of DNA barcoding methods, a summary of the analysis completed, and a discussion of the lessons learned.

Part I: DNA Barcoding Literature Review

Over the last decade the field of DNA barcoding has emerged as a molecular method for species identification. DNA barcoding relies on a uniform region of the mitochondrial gene being amplified, sequenced, and analyzed by comparison to an open access database. Using molecular taxonomy to create a biological barcode that identifies organisms is the central goal of DNA barcoding, as well as creating a standardized reference library for the DNA based identification of target species (Kerr et al. 2007). There are two central principals of DNA barcoding: standardization of the PCR methods and protocols, and the ability to grow the data as the science progresses (Hollingsworth et al. 2011). DNA barcoding can correct field misidentification, reduces ambiguity of species identification, makes species identification more exact, democratizes access by creating open access databases, and expands technical expertise of taxonomists (Stoeckle et al. 2004). The precise identification of organisms has been the realm of taxonomic experts who use specialized language and literature to describe and identify an organism; DNA based identification systems use standardized molecular biology techniques (DNA extraction, PCR, and DNA sequencing) that can increase the speed of the identification of an unknown organism (Seifert et al. 2007). The goal of scientists who perform DNA barcoding is to create a library of every organism on earth (Stoeckle et al.2004).

The purpose of this document is to provide a general background on DNA barcoding, and summarize the methods that are employed by the Reclamation Detection Laboratory for Exotic Species (RDLES). There are many reviews, publications and books available focusing on the methods and protocols involved in DNA barcoding (Lopez and Erickson, 2012). The Canadian Center for DNA Barcoding (CCDB) is a leader in DNA barcoding, with published protocols on a wide range of organisms. CCDB has developed protocols for PCR primers, master mixes, and PCR programs to amplify DNA, which are published online and are straight forward to perform (CCDB website).

DNA Barcoding Work Flow

DNA barcoding has three main steps: DNA extraction, PCR amplification, and DNA sequencing and analysis (Figure 1). DNA isolation is a key step because, without high quality DNA, the PCR amplification will not be optimal. The PCR amplification has to work so that there is DNA for sequencing. And finally, the sequencing analysis has to be successful for there to be an identification of the organism. Ensuring that these three steps are optimal is important for successful DNA barcoding. It is important to note that modifications to the DNA extraction process can sometimes be necessary.

In an effort to streamline DNA barcoding at RDLES, a barcoding standard operating procedure (SOP) was created (Keele et al. 2014). RDLES uses the Qiagen DNA extraction method to extract DNA for barcoding; this method is also used by the FDA for the analysis of fish samples (Handy et al. 2011). This method requires approximately 10 mg of sample tissue.

Universal primer pairs are used to amplify a known region of the cytochrome oxidase I (COI) gene. By amplifying the same gene from diverse organisms it is possible to build a peer-reviewed library of gene sequences. It is important to know the taxonomic group (fish, bird, mammal, etc.) of the organism of interest because the PCR primers are

specific to taxonomic group. For some taxomic groups (plants) genes other than COI are used for DNA barcoding.

Following the PCR amplification, the PCR product is analyzed on an agarose gel to confirm that amplification has occurred. If there is a band, the PCR product can be sent for DNA sequencing. If there is no amplification, it will be necessary to troubleshoot the issue. This might require repeating the DNA extraction, trying out a different primer pair, or changing the master mix.

Once a PCR product has been obtained it is sent to a sequencing company to determine the identity of the organism. The sequencing company provides a ~700 base pair DNA sequence that without bioinformatics has no meaning. There are two programs that can be used to analyze the DNA sequence: Barcode of Life Data Systems (BOLD) and National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST). Each program has positive and negative attributes.

DNA BOLD is a free program that compares the DNA sequence to samples that have been identified by a taxonomist and include additional data about the sample. This website is a hub for DNA barcoding information and analysis. NCBI BLAST is also a free program available on-line, were researchers can submit non-vouchered DNA sequences to the database. This program will compare a sequence to both vouchered and non-vouchered samples. Both of these programs use alignment programs to determine the identity of the unknown sequence. It is helpful to use both independent databases to identify the organism because it can increase the certainty of the identification. It is possible that the two programs will not agree on the identification it and not all organisms are in both the databases.

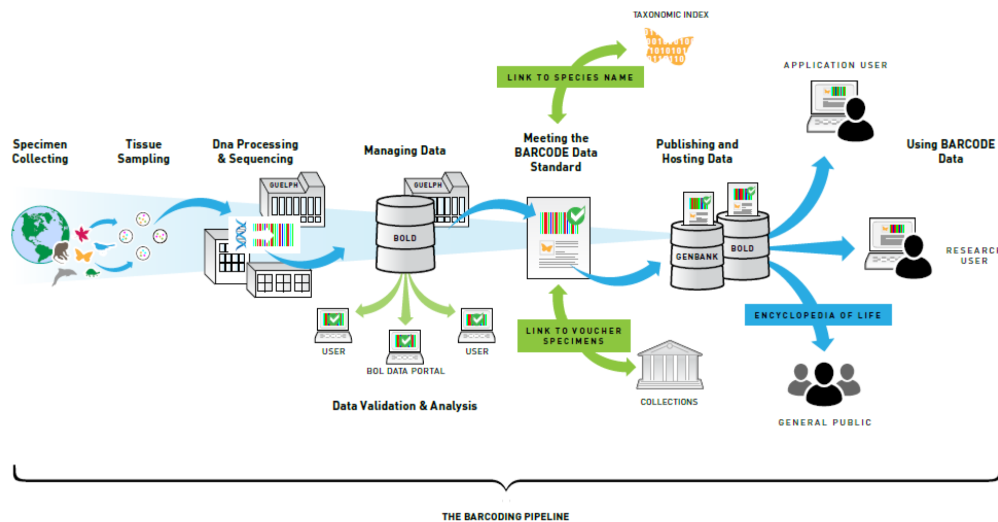


Figure 1: DNA barcoding work flow.
http://www.barcodeoflife.org/sites/all/themes/cbol/pdf/barcode_pipeline.pdf

Specific Taxonomic Groups- Background and Methods

Invertebrates
Folmer et al. (1994) described “universal” DNA primers (named LCO1490 and HCO2198) for a 710 bp fragment of the mitochondrial cytochrome c oxidase I (COI) gene from eleven invertebrate phyla. These phyla include: Echinodermata, Mollusca, Annelida, Pogonophora, Arthropoda, Nemertinea, Echiura, Sipuncula, Platyhelminthes, Tardigrada and Coelenterata, and Vestimentifera. This publication helped to initiate the field of DNA barcoding. The original molecular technique was developed for phylogenetic studies of organisms from deep sea hydrothermal vents and cold water sulfide or methane seep communities. The PCR primers that these researchers designed have since been used for a wide range of studies and have been used to amplify DNA from more than 80 species. Whole cell DNA was isolated using a conventional hexadecyl-trimethyl-ammonium bromide (CTAB) protocol. The PCR was done in a 50 µL reaction mixture with 1 µL of DNA. Following amplification, the PCR product was analyzed on an agarose gel. Once the PCR amplification was performed, sequencing was done to verify that the sequence was COI and to build phylogenic trees.

Fish
There have been multiple papers that use DNA barcoding to identify fish. The FDA is performing these analyses because it is important to ensure that fish available on the food market are labeled correctly by industry.

In 2007, Ivanova et al., proposed the use of a primer cocktail (three forward and three reverse primers with M13 DNA fragments in the PCR primers) to amplify COI from representatives of 94 fish families. In this publication, M13-tailed primers were used to facilitate the sequencing and it was found that by incorporating the M13 tail into the forward and reverse primers it is possible to perform high throughput barcoding on taxonomically diverse samples. Each primer in the cocktail had M13-tails present; this enabled the researchers to use M13 sequencing primers to sequence the PCR products without having to use three different forward and reverse primer pairs.

The FDA has published a detailed SOP online for generating DNA barcodes suitable for species identification of an unknown fish tissue sample based on the Handy et al. (2011) publication. Handy et al. were able to build on the work of Ivanova et al. (2007) to create a single laboratory validated method for the generation of DNA barcodes that would meet regulatory compliance. The FDA’s SOP is robust and easy to follow. This SOP has been used at RDLES for the DNA barcoding of fish (Keele et al. 2014).

Additional publications have surveyed the diversity of North American fish. April et al. (2011) obtained the barcodes for 5,674 fish species (50 families, 178 genera, and 752 species) and was able to obtain sequences for more than 80% of the 902 Canadian and American species listed in the book “Common and Scientific Names of Fishes from the United Stated, Canada, and Mexico” (Nelson 2004). The researchers were able to demonstrate that 90% of the fish sequences could be used to identify the organism by DNA barcoding. These authors also showed that the current fish taxonomy concealed diversity in some of the groups. For example, of the 752 expertly identified museum specimens analyzed, the researchers found 138 samples that needed to be reassessed by taxonomists. This research may help discover increased species diversity of fresh water fish in North America. The authors estimate that as many as 28% of the fresh water fish in Canada and America needed formal taxonomic descriptions. In the future, the use of DNA barcoding will expand as fish populations are threatened and change.

Birds
Extensive DNA barcoding research has also been completed for birds. Hebert et al. (2004) was able to determine the DNA barcodes for 260 species of North American birds. All 260 species of birds had different COI sequences, and the differences between the closely related species was higher (18X) than the differences within a species. The researchers proposed that a 10-fold difference between DNA sequences could be used as a standard screening threshold to determine a new species. By using this threshold Herbert et al. (2004) was able to identity four new species of birds in North America.

Kerr et al. (2007) analyzed 643 species of North American birds primarily using the BirdF1, and BirdR1 primers. If the amplification was not successful then additional primers (FalcoFa, BirdR2, or VertebrateR1) were used. One reason the BirdF1/R1 primers did not always amplify the DNA was the significant difference between the DNA sequence and primer sequence prevented annealing. Most (94%) of the species analyzed had distinct barcodes. In the remaining 6%, the barcode clusters corresponded to small sets of closely related species that are known to hybridize.

Mammals
Amplification of the COI gene for DNA barcoding can be difficult for mammals. This has led to the development of primer cocktails that contain multiple forward and reverse primers that contain the M13 sequence to ensure coverage of the COI gene.

DNA barcoding was used to study Neotropical bats from Guyana (Clare et al. 2007). In this publication, the authors used the glass fiber protocol to isolate DNA from a 1-mm^3 piece of frozen tissue (liver, heart or kidney). The target COI was amplified using two different mammalian barcoding cocktails: C_VF1di and C_VR1di. In addition, an improved primer cocktail that contained M13 tailed versions of the primers (C_VF/C_VR) and an additional primer pair (LepF1_t1 and LepR1_t1) was used to determine the relationships between multiple species of bats.

Mini-barcodes that are approximately 100 bp in size have been designed and used with next generation sequencing to amplify thousands of DNA sequences at once. This approach was used for the analysis of rodent samples (Galan et al. 2012). Researchers designed primers to a 136 base pair fragment of the cytochrome b gene by aligning 9,071 rodent sequences and looking at the conserved region of the gene. The next generation sequencing was able to tag, multiplex, and sequence 1,140 amplicons in a single run. The researchers were able to validate the method on 265 identified rodent tissues that were from 103 different species. Mini-barcodes are short, ~150 base pair fragments, of the COI gene. Because of their size, it is possible to use next generation sequencing which allows for high throughput screening when all of the sample are analyzed in parallel. This research shows the potential of next generation sequencing for obtaining accurate species identification using mini-barcodes. This technology could be applied to a broad range of organisms. This method will enable scientists to increase accuracy and decrease the cost and time need to perform DNA barcoding.

Plants
Plants represent a more complex barcoding problem than other eukaryotes (such as animals) because plant mitochondrial genomes have a low nucleotide substitution rate (Hollingsworth et al. 2011). It has been found that genes other than COI should be used for plant identification because there are not enough changes in the COI between

different plant groups. Various combinations of plant specific markers (*rpoC1+rpoB+matK* or *rpoC1+matK+trnH-psbA*; *rbcL+trnH-psbA*; *atpF-H+psbK-I+matK)* can be used for plant barcoding. The current literature seems to be coming to the conclusion that two or more markers are needed to identify plants. There is still a great deal of research that needs to be done to create markers that can reliability analyze divergent taxonomic groups.

Conclusions

Designing and developing taxonomic specific DNA barcoding methods has been an ongoing process for researchers around the world. Methods are always improving and it is important to remain current with the barcoding literature to identify new primers and methods as new taxonomic groups are analyzed and methods are developed.

Part II: DNA Barcoding at Reclamation

The correct identification of invasive, threatened, and endangered species is important to the management of the facilities and land that Reclamation is responsible for preserving. Over the last ten years, DNA barcoding has increased researchers ability to identify organisms by molecular methods. DNA barcoding can be used by Reclamation researchers on a wide range of projects to identify fish, birds, and insects. Barcoding can help identify all life stages of an organism, which is often difficult to do using traditional taxonomic methods. In addition, the DNA barcoding primers can be used as a starting place in the design of species specific PCR primers. This research was designed to identify ways that DNA barcoding can be used by Reclamation researchers. Several taxonomic groups (fish, birds, and insects) that are of interest to Reclamation were analyzed at RDLES. It is possible for researchers at Reclamation to use DNA barcoding for the detection of invasive and endangered species.

DNA barcoding is performed on monotypic samples. These samples are not like the raw water samples that are currently analyzed at RDLES for quagga and zebra mussel DNA detection. A monotypic sample consists of either the whole, or part of a single organisms body. It is easier to extract DNA and get a PCR product from a monotypic sample than from a complex, mixed environmental sample.

DNA barcoding can also be used as a starting place in the design of species specific primers that can be used for raw water and other environmental samples. For example, the Folmer et al. (1994) COI primers were used by Claxton et al. (1998) as a starting place to design primers that were specific to quagga and zebra mussels. Therefore, a monotypic sample can be analyzed with DNA barcoding primers to produce a DNA sequence that can be used to design new species specific primers, which can in turn be used to analyze a complex environmental sample to detect an organism of interest. This method of primer design is useful because some organisms do not have transcripts available in any database to use as a starting place for designing a molecular assay.

DNA barcoding has been used to differentiate invasive *Dreissena* species in Europe, as multiple species can be present in the same water body (Marescaux et al. 2013). The authors of this publication used the Qiagen DNeasy blood and tissue kit to isolate DNA and then the Folmer et al. primers to amplify COI from 241 *Dreissena* mussels collected from the Meuse River. The sequencing data was analyzed for unique haplotypes and used to construction a haplotype network to determine the relationships between populations.

The authors were able to show that the DNA barcode could be used to identify the two different *Dreissena* species.

Barcoding for the Identification of Invasive and Endangered Species
There are a several invasive organisms that are of interest to Reclamation researchers. These organisms include: Eurasian water milfoil (*Myriophyllum spicatum*), water hyacinth (*Eichornia crassipes*), hydrilla (*Hydrilla verticillata*), purple loosestrife (*Lythrum salicaria*), curly leaf pondweed (*Potamogeton crispus*), didymo (*Didymosphenia geminate*), golden algae (*Chrysochromulina parva Lackey*), Salvinia (*Salvinia spp.*), New Zealand mud snail (*Potamopyrgus antipodarum*), rusty crayfish (*Orconectes rusticus*), and apple snail (*pomacea spp.*). In addition, Reclamation researchers are studying endangered or threatened species (fish, birds, insects) where having the molecular identification will help to confirm the field identification of the organism. As a first step, an analysis of the DNA BOLD and PubMed databases was performed to determine if sequencing for all of the invasive organisms of interest were available. Table 1 provides a summary the available genetic information about several organisms of concern to Reclamation. This table shows that all organisms of interest, except for didymo are present in the DNA BOLD database. There are five nucleotide sequences for didymo in PubMed, and there has been at least one report from New Zealand where molecular methods for the detection of didymo are discussed (Cary et al. 2007). Many of the organisms listed in Table 1 have multiple records in BOLD. PubMed provides additional information about these species, including publications and nucleotide and protein sequences.

In order to identify organisms that are not present in either database it is necessary to look at closely related organisms to get a genus level identification. When Reclamation identifies an organism not previously included in the databases it should be submitted so that other researchers will have access.

Table 1: List of invasive organisms of concern and the number of listed records of DNA sequences from both the DNA BOLD and PubMed databases.

Invasive Organism	**DNA BOLD (Listed Nucleotide Sequences)**	**PubMed (Listed Nucleotide Sequences)**
Eurasian water milfoil (*Myriophyllum spicatum*)	13 sequences	152 sequences
Water hyacinth (*Eichhornia crassipes*)	9 sequences	231 sequences
Hydrilla (*Hydrilla verticillata)*	9 sequences	299 sequences
Purple loosestrife (*Lythrum salicaria)*	5 sequences	69 sequences
Didymo (*Didymosphenia geminate*)	0 sequences	5 sequences
Golden algae (*Chrysochromulina parva Lackey*)	12 sequences	1 sequences
Salvinia (*Salvinia spp.*)	12 sequences	18 sequences
New Zealand mud snail (*Potamopyrgus antipodarum)*	25 sequences	310 sequences
Rusty crayfish (*Orconectes rusticus*)	5 sequences	23 sequences
Apple snail (*Pomacea spp.*)	100 sequences	762 sequences
Quagga Mussel (*Dreissena rostriformis bugensis*)	2 sequence	126 sequences
Zebra Mussel (*Dreissena polymorpha)*	126 sequences	1161 sequence

Creation of a Barcoding Standard Operating Procedure
In an effort to standardize the handling of DNA barcoding samples at RDLES a standard operating procedure (SOP) was created for DNA barcoding (Keele et al. 2014). To summarize the SOP, small tissue samples are taken from the organism of interest and placed into a 1.5-mL eppendorf tube. The DNA is then extracted according to the FDA Fish Barcoding Method (Handy et al. 2011). Once the DNA is extracted the sample is analyzed by PCR. Based on the taxonomic group (invertebrate, plant, mammal, ect.) the primers and the PCR program are selected. In addition, a sample that is known to give a positive band for the primers and PCR program is included as a positive control. A negative control containing no DNA is also used to ensure that the master mix does not contain contaminating DNA. Following the PCR amplification, the PCR product is analyzed by agarose gel electrophoresis and if a band is present at the expected size, the PCR product is sent for DNA sequencing.

Once the DNA sequencing is completed, the resulting sequencing is analyzed using the DNA Bold program and NCBI BLAST program to determine the identity of the organism. This general SOP is useful because it provides a starting place in the analysis of samples. Not every sample can be amplified using this method, as some samples require modifications to the DNA extraction and PCR analysis. These modifications are determined on a case by case basis.

Fish Barcoding

Fin clips from 10 fish species were analyzed at RIDLES to test the effectiveness of barcoding primers and methods. Each of the species analyzed produced a positive PCR band (Figure 2). It is important to note that all of the bands look exactly alike. The only way to determine the identity of the organism is through DNA sequencing. The COI primers are designed to give a band that is approximately 700 base pairs in size.

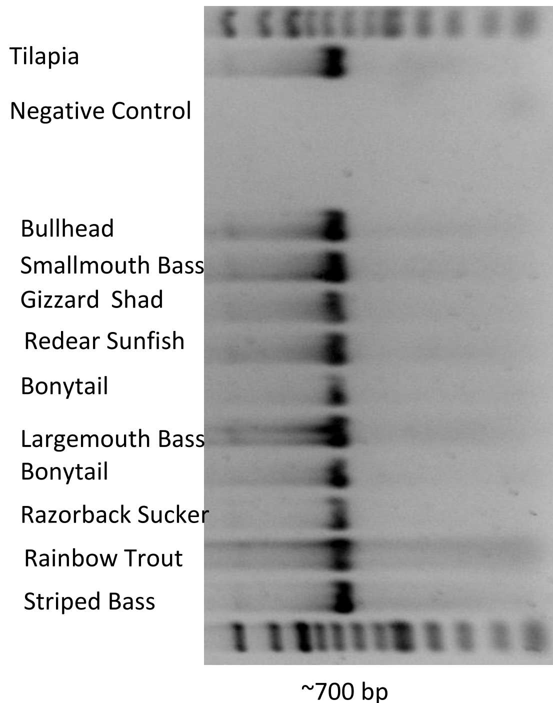


Figure 2: Gel electrophoresis results of DNA barcoding from fin clips from10 species of fish.

Both the DNA Bold and PubMed analysis confirmed that the field identification of the striped bass, razorback sucker, bonytail chub, largemouth bass and yellow bullhead samples. The redear sunfish sample was identified correctly in PubMed, but the BOLD databases identified it as either a warmouth sunfish or redear sunfish. The gizzard shad was identified as an American gizzard shad in Pubmed and in BOLD as an American gizzard shad, Mexican gizzard shad, or Menhaden.

The fish identified as a smallmouth bass in the field, was identified as a largemouth bass via barcoding. It is likely that this fish was a smallmouth and largemouth hybrid, with smallmouth bass phenotypic characteristics. The COI is inherited along the maternal line; therefore, if the father had been a smallmouth bass it could not be determined with this test.

A species level identification of the tilapia positive control could not be made with the BOLD program. Several different species of tilapia were suggested by the BOLD program, and these results were not surprising, considering the specimen purchased from the grocery store was likely a farm raised fish with a mixed genetic background.

Lamprey Samples
Nine lamprey samples from California were submitted for DNA barcoding by a Reclamation researcher. All nine samples produced enough DNA to be sequenced and all sequences were correctly identified as *Entosphenus hubbsi* (Kern Brook Lamprey). These results confirmed the field identification of these organisms.

Two Fish Samples from California
Two fish fin clips collected in California were analyzed for a researcher at Reclamation, and were found to match the field identification. Sample one was an *Oncorhynchus mykiss* (rainbow trout), and sample two was an *Oncorhynchus keta* (chum salmon). The results for the chum salmon are interesting because the field and molecular identification place this fish in an area that was previously thought to be outside its normal range. This finding prompted the Reclamation researcher to write a short note about the unique location of this collection that will be submitted for publication.

Twenty Fish Samples from Divide Creek
Twenty fish fin clips collected on July 31, 2013 from Divide Creek were analyzed and found to be *Oncorhynchus clarkii* (cutthroat trout) with 99-100% certainty. Both the BOLD and NCBI BLAST programs give *Oncorhynchus* as the genus. BOLD would not differentiate between *O. clarki* and *O. clarkii*. The BLAST program predicted that the organism is *O. clarkii*.

Overall, four different fish barcoding projects have been successful. Fish barcoding can be offered as a service at RDLES for researchers who want to have a molecular identification for the samples that they collect. This method can be used to confirm adult identifications as well as identify fish larvae.

Bird Barcoding
In order to test the bird specific PCR primers and PCR method, two bird feather samples, one from a pet bird and the second from a back-yard birdfeeder were analyzed with bird specific COI PCR primers. The sequencing identified the pet bird as either *Streptopelia roseogrisea* (African collared dove) or as *Streptopelia capicola* (ring necked dove). The bird had been sold as a Japanese moon dove (no scientific name found). There are two possible explanations for the differences in the molecular and visual identification. First, the paternal line could be a moon dove and this bird is a hybrid that had dominant parental traits. Or it is possible that the seller calls all of these types of birds moon doves.

The second bird sample was collected from a back yard feeder and was identified as a *Pica hudsonia* (black billed magpie). The alignment of this sequence was very close to *Pica nuttalli* (yellow billed magpie). There are only two nucleotide differences between the two sequences. This leads to an important question for DNA barcoding analysis: what are the cutoffs for delineating between different species? This will be an ongoing issue that will have to be considered when performing DNA barcoding analysis.

Additional bird feather samples collected by Reclamation field biologists were analyzed to determine the molecular identification of their samples. It was more difficult to isolate DNA from feather samples than fish fin clips. It was necessary to increase the amount of reagents and time used in the DNA extraction to get quality DNA for the PCR reaction. Once the PCR was performed some of the bird samples produced a larger band (~1200 bp) than the expected (~700 bp) band (Figure 3).The reverse sequence identified the birds as *Empidonax trailli* (willow flycatcher). All seven of the samples analyzed contained the

same nucleotide differences from the voucher specimen (Figure 4). The analysis of the forward sequence in the DNA BOLD program did not provide an identity (Figure 5).

The NCBI BLAST program analysis showed that these sequences all had ~80% identity with a flycatcher species. Based on these results it is not possible to use the forward reaction to identity the organism. Sequence alignment with the *E. trailli* voucher sequence and the reverse sequence indicates that the forward sequence overlaps the voucher sequence for ~140 base pairs and the majority of the forward sequence does not align to the voucher sequence. The reverse sequence aligns to the end of the voucher sequence. The middle of the voucher sequence has no coverage with the sequenced DNA. It is possible that the forward primer is setting down upstream of the expected site, and thus producing the larger band.

Although all seven samples were identified at *E. trailli*, the sequence alignment is not 100% perfect. All seven samples align with each other, but there are clear differences when compared to the voucher sequence. This finding leads to an important question. How many nucleotide differences are needed for an organism to be considered a separate species or subspecies? Are the differences seen here due to these birds being a separate sub-species? More samples will have to be analyzed to determine if these nucleotide differences are consistent in this bird population. Also, additional flycatchers should be analyzed to build a clearer picture of the genus.

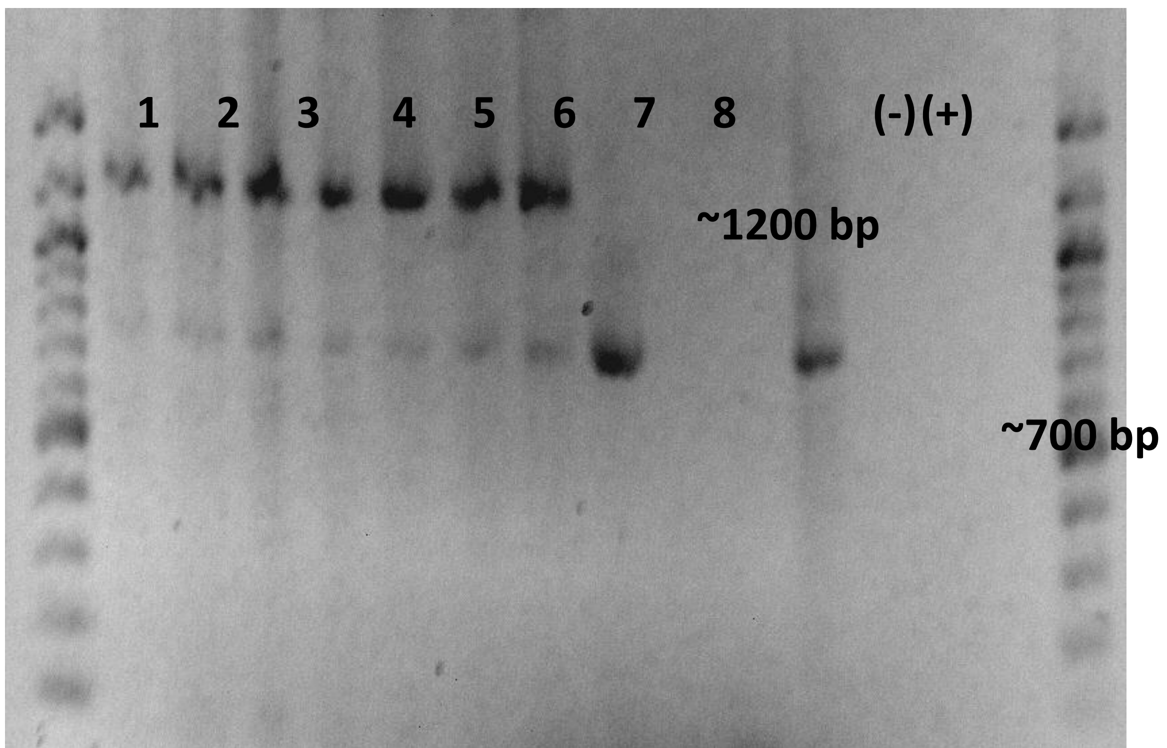


Figure 3: Gel electrophoresis analysis of the PCR products from eight bird samples 1 through 7 shows the double banding effect. The size of the PCR product should be ~700 bp; instead there is s band at ~1200 bp for these seven samples. Sample 8 and the positive control yielded the band of the correct size.

```
Thu Jul 24, 2014  9:11 -0600
E trailli voucher BIOUG FLAP 5551.str from 585 to 1
Alignment to
Bird 1 reverse.str--  Matches:184; Mismatches:4; Gaps:405; Unattempted:0
Bird 2 reverse.str--  Matches:194; Mismatches:4; Gaps:387; Unattempted:0
Bird 3 reverse.str--  Matches:226; Mismatches:11; Gaps:354; Unattempted:0
Bird 4 reverse.str--  Matches:539; Mismatches:39; Gaps:18; Unattempted:0
Bird 6 reverse.str--  Matches:194; Mismatches:7; Gaps:395; Unattempted:0
Bird 7 reverse.str--  Matches:390; Mismatches:17; Gaps:181; Unattempted:0

                *         *         *         *         *         *         *         *         *
585<g----a---gatgctgatacagcactgggtctccaccccctgcggggtcaaagaatgtagtgttaaggttacggtctgttaatagcatagtgataccggc<493
192<GAANNA---GATGTTGATACAGCACTGGGTCTCCACCCCCTGCGGGATCAAAGAATGTAGTGTTAAGGTTACGGTCTGTTAATAGCATCGTGATACCGGC<96
198<~~~~~~~~~~~t--tggtacagcactgggtctccaccccctgcgggatcaaagaatgtagtgttaaggttacggtctgttaatagcatggtgataccggc<112
243<~~~~~~~~~~~~~~~~ATACAGCACTGGGTCTCCACCCCCTGCGGGATCAAAGAATGTAGTGTTAAGGTTACGGTCTGTTAATAGCATCGTGATACCGGC<160
578<G----A---GATGTTGATACAGCACTGGGTCTCCACCCCCTGCGGGATCAAAGAATGTAGTGTTAAGGTTACGGTCTGTTAATAGCATCGTGATACCGGC<486
 12>G----A---GANGTTGCTACAGCACTGGGTCTCCACCCCCTGCGGGATCAAAGAATGTAGTGTTAAGGTTACGGTCTGTTAATAGCATCGTGATACCGGC>104
410<G----AAGAGATGTTGATACAGCACTGGGTCTCCACCCCCTGCGGGATCAAAGAATGTAGTGTTAAGGTTACGGTCTGTTAATAGCATCGTGATACCGGC<315

      *         *         *         *         *         *         *         *         *         *
492<agcaagaactggtagagagaggaggagaagaactgcggtaattaggacggatcatacaaataagggagtttggtattgtgatagagcggggggttttata<393
 95<AGCGAGAACTGGTAGAGAGAGGAGGAGAAGAACTGCGGTAATTAGGACGGATCATACAAATAAGGGAGTTTGGTATTGTGATAGAGCGGGGGGTT~~~~~<1
111<agcgagaactggtagagagaggaggagaagaactgcggtaattaggacggatcatacaaataagggagtttggtattgtgatagagcggggggttt----<16
159<AGCGAGAACTGGTAGAGAGAGGAGGAGAAGAACTGCGGTAATTAGGACGGATCATACAAATAAGGGAGTTTGGTATTGTGATAGAGCGGGGGGTTT----<64
485<AGCGAGAACTGGTAGAGAGAGGAGGAGAAGAACTGCGGTAATTAGGACGGATCATACAAATAAGGGAGTTTGGTATTGTGATAGAGCGGGGGGTTTCATA<386
105>AGCGAGAACTGGTAGAGAGAGGAGGAGAAGAACTGCGGTAATTAGGACGGATCATACAAATAAGGGAGTTTGGTATTGTGATAGAGCGGGGGGTTT---->200
314<AGCGAGAACTGGTAGAGAGAGGAGGAGAAGAACTGCGGTAATTAGGACGGATCATACAAATAAGGGAGTTTGGTATTGTGATAGAGCGGGGGGTTTCATA<215
```

Figure 4: Alignment of the reverse sequence of seven bird samples to the *E. trailli* sequence. The sites that are highlighted in red are different from the voucher specimen. These differences are consistent with all seven birds analyzed.

```
Thu Jul 24, 2014  8:59 -0600
E trailli voucher BIOUG FLAP 5551.str from 1 to 585
Alignment to
Bird 1 Forward.str--  Matches:130; Mismatches:8; Gaps:988; Unattempted:0
Bird 1 reverse.str--  Matches:184; Mismatches:4; Gaps:405; Unattempted:0

             *         *         *         *         *         *         *         *         *         *
  1>gaacttggacaaccaggaactctcctaggagatgaccaaatttacaacgtaatcgttactgctcatgccttcgtaataatcttctttatagtaatgccta>100
542>GAACTTGGACAACCAGGAACTCTCCTAGGAGATGACCAAATTTACAACGTNNTCGTTACNGNTCANGCCTTCGTAATAATCTTCTTTATNGTAATGCCTA>641
  0>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>0

             *         *         *         *         *         *         *         *         *         *
101>tcataattggaggatttggtaactgattagttcctctaataattggtgcccctgatatagcattccctcgcataaataacataagcttctgacttctgcc>200
642>TNATAATTGNAGGATTTGGTAACTGATTAGTTCCTCTA~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>679
  0>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>0

             *         *         *         *         *         *         *         *         *         *
201>cccatcatttctcctccttttagcctcctctacagtcgaagccggtgcagggaccggatgaactgtttatccaccattagctggtaatctagcacatgct>300
679>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>679
  0>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>0

             *         *         *         *         *         *         *         *         *         *
301>ggagcttcagtagacttggccattttctctcttcatttagcaggtgtctcctctatcctaggggctattaacttcattactaccgcaattaatataaaac>400
679>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>679
  1>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~AAC>3

             *         *         *         *         *         *         *         *         *         *
401>cccccgctctatcacaataccaaactcccttatttgtatgatccgtcctaattaccgcagttcttctcctcctctctctaccagttcttgctgccggtat>500
679>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>679
  4>CCCCCGCTCTATCACAATACCAAACTCCCTTATTTGTATGATCCGTCCTAATTACCGCAGTTCTTCTCCTCCTCTCTCTACCAGTTCTCGCTGCCGGTAT>103

             *         *         *         *         *         *         *         *
501>cactatgctattaacagaccgtaaccttaacactacattctttgaccccgcagggggtggagacccagtgctgtatcagcatc----tc>585
679>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>679
104>CACCATGCTATTAACAGACCGTAACCTTAACACTACATTCTTTGATCCCGCAGGGGGTGGAGACCCAGTGCTGTATCAACATCTNNTTC>192
```

Figure 5: Alignment of sequences from bird sample 1 with the forward and reverse sequences, note the gap between the bird1 forward and reverse sequences when aligned to the voucher sequence. These results lead to the hypothesis that in the upstream sequence from where the forward primer was supposed to bind, that there was sequence that was similar enough for the forward primer to bind to and start priming.

Insect Barcoding

MacNeill's Saltbush Sootywing (*Hesperopsis gracielae*) Identification by DNA Analysis

The goal of this project was to develop sootywing specific primers that can be used to analyze environmental DNA samples to detect the MacNeill's Saltbush Sootywing (*Hesperopsis gracielae*). Collection of environmental samples provides a non-lethal collection method that helps conserve the threatened species. Voucher sootywing samples were analyzed using the universal DNA primers (Folmer et al. 1994) to obtain an initial COI sequence that could be used to design sootywing specific PCR primers. One of the samples was identified as the moth, *Trichocosmia inornata*, and two of the samples were identified as *Hesperopsis gracielae* (MacNeill's Sootywing) by the DNA BOLD database.

Following the DNA sequencing of the COI gene, the sootywing samples were aligned and primers specific to this organism were designed using IDT DNA Primer Quest. Primer set, SW1, gives a 200 bp PCR product. Primer set, SW2, gives a 210 bp PCR product. Both of these primers were tested against the moth and the two known sootywing samples. The moth sample gave no PCR product while both the sootywing samples gave positive PCR results. The master mix and PCR program used for quagga mussel COI analysis was used with the SW1 and SW2 PCR primers. Additional optimization may be required to make sure the PCR master mix and programs are optimal for this analysis. With these primers it will be possible to analyzed complex samples for the presence of sootywing DNA that would indicate the presence or absence of this organism at a location.

After the sootywing primers were designed, two sootywing egg samples were analyzed with the sootywing specific primers and both produced positive PCR results. These results show that it is possible to obtain a positive PCR result from a small amount of tissue with the sootywing primers.

The next step was to use the primers on environmental samples to determine if sootywing DNA could be detected to determine presence and absence. Environmental samples were collected by placing brightly colored pieces of cloth or paper in the butterfly's habitat to attract the sootywing. In theory, sootywings that landed on the targets left cells behind that could be analyzed by DNA barcoding.

Preliminary results indicate that paper will serve as a better target than cloth. A captured butterfly was placed onto the paper that gave a positive PCR result. This project is ongoing and has the potential to improve the delectability of these butterflies. These experiments show the potential of DNA barcoding, and how it can be used to design species specific PCR primers that can be used to analyze environmental samples.

Conclusions

In addition to fish, birds, and butterflies; mammals were analyzed to test the DNA barcoding primers and methods. Human hair was used to test mammal primers. The PCR was successful and the sequence data showed that the samples were taken from *Homo sapiens*. Plant samples have not yet been tested as they are more difficult to analyze because there are several different PCR primers that can be used and it is difficult to determine the correct primers and PCR program to use. Further research is required, and will continue as analysis of plant samples is requested.

The next steps for this technology will be to expand from the analysis of a single monotypic sample into more complex samples. Mini-barcode analysis of all organisms in a sample, with the use of next generation sequencing technology, will enable researchers to explore the complete ecology of a sample. For example, next generation sequencing of fish gut content can be used to determine the diet of a fish. Fish gut content analysis by mini-barcode analysis is a project that is currently being designed and will be proposed by researchers at Reclamation.

Using DNA barcoding methods at RDLES has involved obtaining the primers and reagents for PCR, performing DNA extractions and PCR amplification on samples, and sequencing the resulting PCR product. If the sample did not amplify then it has been necessary to do troubleshooting on the DNA extraction method, PCR master mix components, and PCR thermocycler program. Overcoming and understanding the issues associated with DNA barcoding has been part of the development process of this technique for researchers at RDLES. This technology validates the field identification of endangered or invasive species. It can also be a starting place for the design of new PCR assays to look at environmental samples.

Part III: Lessons Learned at RDLES about DNA Barcoding

Issue 1: DNA Extraction

The bird feathers did not provid consistent amplification of the PCR product. Different extraction methods (increased reagent amount and time) with the Qiagen blood and tissue kit were used with success to obtain PCR results. These samples illustrated the importance of flexibility in the DNA extraction method.

The current method used for DNA extraction, based on the FDA Fish Barcoding SOP, is one that can be modified to fit the needs of the sample. The volumes of reagents used and the amount of time for incubations can be increased or decreased to ensure that the best quality DNA is extracted. As different samples, such a bird eggs and shells, are analyzed it is likely that the DNA extraction methods will require small changes to obtain the quality of DNA needed for the PCR reaction.

Issue 2: Organism Not Present in any Database

Some of the organisms sampled were not in either the DNA BOLD or PubMed databases. The alignment of these samples was below 90% which implies that the organism in not in any database at this time. It may be possible to assign these organisms to a genus, but a species level identification is impossible. For these samples, it will be necessary to work with a taxonomist for a final identification. Also, it would be important to hopefully be able to create a voucher that could be placed in the DNA BOLD database.

Issue 3: Primers

The Folmer et al. (1994) COI primers can be used for a wide range of organisms (mainly invertebrates). For example, the Folmer et al. primers were used to amplify the COI gene from adult quagga mussel DNA. Two different PCR programs were used; the FDA fish barcoding and RDLES QM PCR method, and both methods worked with the Folmer et al. primers. These results indicate that these primers can be used with different PCR master mix and program settings.

On the other hand, when the Folmer et al. primers were used to analyze a drain sludge sample for cyanobacteria and algae, no bands were detected from the five DNA extracts. Following this initial analysis it was necessary to order primers specific to cyanobacteria. These primers did not work as expected. They did not amplify any cyanobacteria. The sequence analysis did not give a clear identity on the samples, because the PubMed identities were all below 90%. Further research and analysis is needed to determine the best primers to use for algae and cyanobacteria.

Primer selection is important for obtaining a DNA band that can be analyzed. When the primers produce bands that are not of the expected size, the results are difficult to understand. For example, in the bird sample mentioned above, the PCR product was not the expected size (~700 bp), but was instead (~1200 bp) or almost double the size that was expected. In this case, the samples were analyzed twice and both times this larger band was produced. The analysis also produced a faint band at ~700 bp, but that the (~1200 bp) band was more prominent (Figure 3).

The results of the sequencing shed light on what was happing with these samples. The forward PCR primer was actually setting approximately 500 bp upstream of the site where it should be annealing. The forward primer was able to recognize two sites and thus produced the double band. This result has only been seen in this group of birds at RLDES. It will be very interesting to analyze additional samples of this bird species to determine is the double band is a onetime event or will be an phenomenon seen with all samples from this particular species. Nonspecific PCR priming can be a major issue when performing DNA barcoding.

Issue 4: Hybrids
A major issue with DNA barcoding is that it does not account for hybrids. Hybrids occur when two different species mate. If the offspring is analyzed with DNA barcoding only the maternal species would be revealed. The gene that has been selected for DNA barcoding (COI) is inherited through the maternal lineage. Thus, the paternal lineage is masked. If the taxonomist identifies the organism based on dominant paternal traits, the laboratory will only identify the organism based on the material gene. This leads to a conflict in the identification between the field and laboratory. This issue was seen when a researcher identified a sample as a smallmouth bass and the barcoding results showed that it was a largemouth bass. One way to overcome this issue would be to use species specific primers to determine the paternal lineage after the barcoding analysis has been completed.

Summary
DNA barcoding is a useful molecular technique for the identification of unknown organisms at any life stage. This technology can be used on any life stage of an organism and on a wide range of organisms. With this molecular technique it is possible to analyze tissue from seeds or embryos to the fully grown adult organism. This technology decreases the number of organisms that have to be collected in the field and reduces the amount of time between collection and identification. Barcoding can be used as a starting place in the design of species specific primers and assays from environmental samples. Overcoming and understanding the issues associated with barcoding will be an ongoing process as more samples are analyzed by DNA barcoding at RDLES. Over the last year, RLDES has shown that it has the capabilities to analyzed barcoding samples from a wide range of organisms. This new technique will benefit a variety of Reclamation projects by providing fast and cost effective species identification.

Literature Cited

April J, Mayden RL, Hanner RH, Bernatchex L. (2011). Genetic calibration of species diversity among North America's freshwater fishes. *PNAS*, 108(26), 10602-10607.

Armstrong KF, Ball SL. (2005). DNA barcodes for biosecurity: invasive species identification. *Phil. Trans. R. Soc. B*, 360, 1813-1823.

Cary SC, Hicks BJ, Coyne KJ, Ruecket A, Gemmill CEC, Barnett CME. (2007). A sensitive genetic-based detection capability for *Didymosphenia geminate* (Lyngbye) M. Schmidt: Phases Two and Three. *CBER Contract Report 62*, Retrieved from: http://www.biosecurity.govt.nz/files/pests/didymo/didymo-dna-detection-phase-2-3-oct-07.pdf.

Clare EL, Lim BK, Engstrom MD, Eger JL, Hebert PDN. (2007) DNA barcoding of Neotropical bats: species identification and discovery within Guyana. *Molecular Ecology Notes*, 7, 184-190.

Claxton WT and Boulding EG. (1998). A new molecular technique for identifying field collections of zebra mussel (*Dreissena polymorph*a*)* and quagga mussel (*Dreissena bugensis*) veliger larvae applied to eastern Lake Erie, Lake Ontario, and Lake Simcoe. *Can. J. Zool.* 76, 194-198.

Folmer O, Black M, Hoeh W, Lutz R, Vrijemnhoek R. (1994) DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology,* 3(5) 294-299.

Galan M, Pages M, Cosson JF. (2012). Next-generation sequencing for rodent barcoding: species identification from fresh, degraded, and environmental samples. *Plos One*, 7(11), e48374-e48374.

Handy SM, Deeds JR, Ivanova NV, Hebert PD, Hanner RH, Ormos A, Weigt LA, Moore MM, Yancy HF. (2011). A single laboratory validated method for the generation of DNA barcodes for the identification of fish for regulatory compliance. *J. AOAC* 94(1): 201-210.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. (2004). Identification of birds through DNA barcodes. *PLoS Biology*, 2 (10), 1657-1633.

Hollingsworth PM, Graham SW, Little DP. (2011). Choosing and using a plant DNA barcode. *Plos One*, 6 (5), e19254-e19254.

Ivanova NV, Zemlak TS, Hanner RH, Herbert PDN. (2007). Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes*, 1-5.

Keele JA, Carmon J, Hosler D. (2014). DNA barcoding for genetic identification of organisms, DNA barcoding standard operating procedure. Technical Memorandum No. 86-68220-14-08.

Kerr KCR, Stoeckle MY, Dove CJ, Weigt LA, Frances CM, Herbert PDN. (2007). Comprehensive DNA barcode coverage of North American birds. *Molecular Ecology Notes*, 7, 535-543.

Lopez I and Erickson DL. (2012) DNA Barcodes: Methods and Protocols. Humana Press.

Marescaux J and KV Doninck. (2013) "Using DNA barcoding to differentiate invasive *Dreissena* species (Mollusca, Bivalvia)." ZooKeys 365: 235-244.

Nelson JS. (2004) Common and Scientific Names of Fishes from the United States, Canada, and Mexico. American Fisheries Society, 6th Edition.

Seifert KA, Samson RA, Dewaard JR, Houbraken J, Levesque CA, Moncalvo JM, Louis-Seize G, Herbert PD. (2007). Prospects for fungus identification using *COI* DNA barcodes, with *Penicillium* as a test case study. PNAS, 104(10),3901-3906

Stoeckle M, Waggoner PE, Ausubel JS. (2004). Barcoding Life: Ten Reasons. Consortium for the Barcode of Life, v3.0. Retrieved from: http://www.barcodeoflife.org/content/barcoding-life-ten-reasons-pamphlet.

Share drive folder, name and path where data are stored:

H:/EnvRes Share/Mussel Samples/2014 Research Proposals/DNABarcoding

Points of Contact:

Jacque Keele:jkeele@usbr.gov (303-445-2187)
Denise Hosler: dhosler@usbr.gov (303-445-2195)
Sherri Pucherelli: spucherelli@usbr.gov (303-445-2015)