Chapter 6

DNA Barcoding Fishes

Lee A. Weigt, Amy C. Driskell, Carole C. Baldwin, and Andrea Ormos

Abstract

This chapter is an overview of the techniques for DNA barcoding of fishes from field collection to DNA sequence analysis. Recommendations for modifications of field protocols and best tissue sampling practices are made. A variety of DNA extraction protocols is provided, including high-throughput robot-assisted methods. A pair of well-tested forward and reverse primers for PCR amplification and sequencing are presented. These primers have been successfully used for DNA barcode on a wide array of marine fish taxa and also work well in most freshwater and cartilaginous fishes. Recipes and cycling protocols for both PCR amplification and sequencing and cleanup methods for the reaction products are provided. A method for the consistent production of high-quality DNA barcodes from DNA sequence data is given and stringent guidelines for judging the quality of raw sequence data are laid out.

Key words: DNA barcoding, Fish, PCR, DNA extraction

1. Introduction

Fishes are the largest and most diverse class of vertebrates and fortunately among the easiest groups for which to generate DNA barcode data. Estimates of species numbers generally exceed 30,000 (1), with about 300 new species described each year. Commercial fisheries value estimates exceed US$200 billion, and efforts in our lab have contributed to standardized protocols for fish barcoding (2, 3), other standardized field surveys and bioassessments (4–6), and testing standardized storage methods and ethanol-recycling instrumentation for contaminant carryover (7, 8).

Eggs, larvae, and juvenile fish can be difficult to identify to species using morphology alone, and we have promoted DNA barcoding as a tool to assist all these efforts (9–11). We find that the correct taxonomy can be illuminated in complex and morphologically challenging species groups by combining comparative investigation of living color patterns, examination of traditional

W. John Kress and David L. Erickson (eds.), *DNA Barcodes: Methods and Protocols*, Methods in Molecular Biology, vol. 858,
DOI 10.1007/978-1-61779-591-6_6,

morphological characters in preserved specimens, and DNA barcodes (12, 13).

Efforts in our lab have focused on bony marine fish, but we have used these methods on both freshwater and cartilaginous fish, though there are other larger barcoding campaigns for both groups (14). We have focused on full-length barcode sequences, but, based on the success of others (15), will be starting "mini-barcode" work shortly on major collections of scientific importance collected 50+ years ago.

We have found high-resolution digital images of freshly captured specimens to be extremely useful in elucidating species in taxonomically challenging species groups, and each of our specimens is photographed as soon as possible after collection and before preservation. Though it is possible to work with many types of fish tissue including scales, fin clips, blood, buccal swabs, bones, and others (16, 17), our preferred starting material is a small muscle biopsy from a fresh specimen. Since we photograph the fish (and subsequently examine much of the morphology) from the left side (fish "looking" to the left, in lateral view), we remove the biopsy from the right side and try to minimize damage to any important morphological characters.

DNA extraction protocols vary by lab, but with fresh tissue from fish, almost all extraction protocols will yield amplifiable DNA. Significant value can be added to a specimen by performing archival quality DNA extractions, though these are more expensive than some alternative extraction methods (e.g., Chelex) that do not produce extracts viable for a great length of time. Some newer extraction protocols have emerged from the Guelph Barcoding lab (18), but as they are newer, the extracts have not been tested for longevity in archival biobanks. The "gold standard" for DNA extractions from animals is the phenol:chloroform method (19), but if performed manually, it has some critical drawbacks, including handling and disposing of hazardous chemicals. We present here our primary method—an automated version of the phenol–chloroform protocol that reduces the problematic factors. However, as not everyone will be able to afford the instrument on which this protocol is performed, we also present a manual method and other alternatives. These include a smaller instrument which utilizes a magnetic bead-based approach and a common filter membrane-based kit.

Amplification of fish DNA via PCR is rarely problematic, and follows standard protocols. The primers are very robust, amplifying almost all taxa tested to date. The same is true for DNA sequencing. Finally, as much of fish biodiversity is known, we have been able to generate a reference sequence. This reference (with degenerate bases to represent variation among fish taxa) eases the alignment and analysis of fish data.

We list specific brands used in our lab, but acceptable equivalents usually exist, for both consumables and many instruments.

2. Materials

Materials common to all laboratory steps include latex/nitrile gloves, as well as pipettes with disposable tips, both filtered and nonfiltered. In addition, use of a centrifuge with a rotor that can accommodate a microtiter plate and is capable of speeds greater than 1,000 rcf, and a thermal cycler, is required for many steps.

2.1. Sample Collection

1. Photo documentation: Digital camera, scale bar, color scale.
2. Tissue sampling: Scalpels, tweezers, tubes (Matrix-brand 2D labeled or other Matrix tubes or 2.0-ml cryovials), handheld barcode scanner, data-recording materials (computer with spreadsheet when possible), bleach or Decon ELIMINase, alcohol burner.

2.2. Tissue Storage

1. Tissue preservation buffer (21): 0.25 M EDTA, 25% DMSO, saturated with NaCl—500 ml DMSO, 1 L 0.5 M EDTA pH 8.0, 500 ml water, >200 g NaCl. Stir solution while adding salt; continue adding salt until no more goes into solution and it begins to collect on the bottom of the mixing vessel.
2. 95% Ethanol (see Note 1).

2.3. DNA Extraction: Automated and Manual Extractions

1. Autogen Prep 965 DNA extraction: Autogen 965 robot, and kit buffers including M1, M2, R3, R4, R5, R6, R7, R8, R9, and proteinase-K for use with animal extractions; 96-well deep well plates (Costar #3960).
2. Automated extractions using Qiagen BioSprint96: Biosprint robot, and BioSprint 96 DNA Blood Kit (940057), Buffer ATL (Qiagen 19076), and proteinase-K.
3. For both automated extraction methods: AxyMat silicone lids (Axygen) for 96-well digestion blocks, plexiglass (or other firm solid material) rectangles cut to fit the tops of the 96-well blocks, and 0.2% Tween® 20: 200 µl Tween® 20 in 100 ml H_2O.
4. Manual extraction lysis buffer: 100 mM EDTA, 25 mM Tris pH 7.5, and 1% SDS.
5. 100 µg/ml proteinase-K dissolved in water.
6. Phenol, equilibrated to pH 7.5 with Tris–HCl pH 8.0.
7. Chloroform:isoamyl alcohol (24:1 ratio).
8. TE solution: 10 mM Tris, 1 mM EDTA pH 8.0; equilibrated to pH 7.6 with Tris–HCl pH 7.0.
9. Incubator or incubator shaker for tissue digestion, capable of maintaining a temperature of >50°C.

2.4. Polymerase Chain Reaction: Amplification and Purification

1. 10 mM deoxynucleotide (dNTP) mix.
2. 100 μM oligonucleotide Primers (IDT Technologies, USA).
3. Biolase Taq DNA Polymerase (BioLine).
4. 10× PCR Buffer for Bioline Taq.
5. 50 mM Magnesium chloride.
6. Liquidator 96-channel benchtop pipette (Rainin).
7. ExoSAP-IT (USB 78201) for purification.
8. 96-Well (0.2 μl volume) plastic PCR plates (Genemate T-3060-1).
9. Silicone plate mat (lid) for PCR plates (Genemate T-3161-1).

2.5. Polymerase Chain Reaction: Visualization

1. Agarose.
2. 1× TBE buffer: 0.9 M Tris base, 0.89 M boric acid, 0.02 M Na-EDTA; prepare by mixing 108 g Tris, 55 g boric acid, and 7.4 g Na-EDTA in a beaker together with 400 ml water, mix until dissolved, and add deionized water to 10 L.
3. Sample loading dye: 0.083% bromophenol blue; 0.083% xylene cyanol, and 10% glycerol.
4. DNA stain: Ethidium bromide (10 mg/ml) or SYBR SAFE (Invitrogen).
5. Optional: DNA size standard ("ladder"; Hi Lo DNA marker, Minnesota Molecular, Inc.).
6. Electrophoresis rig and power supply.
7. Gel imaging system/camera for use over UV light box.

2.6. Sanger Sequencing Components: BigDye Reactions

1. 5× Sequencing Buffer: 400 nm Tris–HCl pH 9.0, and 10 mM $MgCl_2$.
2. Oligonucleotide Primers: Dissolved in water to 10 μM (Table 1 lists all primers used for fish).
3. BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems).
4. 96-Well (0.2 μl volume) plastic PCR plates (Genemate T-3060-1).
5. Silicone plate mat (lid) for PCR plates (Genemate T-3161-1).

2.7. Sephadex Purification of Cycle-Sequencing Products

1. Sephadex® G50 (Sigma).
2. Hi-Di™ formamide (Applied Biosystems).
3. Multiscreen® HTS filter plates (Millipore MSHVN4550).
4. Multiscreen column loader (Millipore MACL09645).
5. Liquidator 96-chanel manual pipette (Rainin).
6. 96-Well (0.2 μl volume) semi-skirted plastic PCR plates (Genemate T-3085-1).
7. Septaseal rubber mats (ABI #4315933).

Table 1
Primer table for fish PCR (and M13 sequencing primers) (from refs. 10, 22, 23)

Barcode primer name	Barcode primer sequence 5′ → 3′
FISHCO1LBC	TCAACYAATCAYAAAGATATYGGCAC
FISHCO1HBC	ACTTCYGGGTGRCCRAARAATCA
FISHCO1LBCm13F	CACGACGTTGTAAAACGACTCAACYAATCAYAAAGATATYGGCAC
FISHCO1HBCm13R	GGATAACAATTTCACACAGGACTTCYGGGTGRCCRAARAATCA
16SAR	CGCCTGTTTATCAAAAACAT
16SBR	CCGGTCTGAACTCAGATCACGT
m13F	CACGACGTTGTAAAACGAC
m13R	GGATAACAATTTCACACAGG

2.8. Genetic Analyzer Components

1. ABI 3130XL genetic analyzer: Polymer POP-7; 36-cm capillary array run using the ABI Template Protocol "RapidSeq36_POP7" with a run time of 2,280 s.
2. ABI 3730XL genetic analyzer: Polymer POP-7; 50-cm capillary array run using the ABI Template Protocol "LongSeq50_POP7" with a run time of 4,000 s.

2.9. Data Processing and Quality Control

1. Sequencher vers. 4.10.1 (Gene Codes).
2. Geneious (BioMatters)—use of the Geneious program and the BioCode plugin is discussed elsewhere in this volume.

3. Methods

3.1. Tissue Sampling

1. *Photography processing*—fish orientation—left side, fish's head on left; see Note 2.
2. *Tissues to sample* (in order of decreasing desirability)—muscle biopsy from right side, right eye, portion of right pectoral fin, other fin clip, gill tissue, swabs, and scales.
3. *Muscle biopsy*—from right side, caudal region, dorsal to lateral line; avoid heavily parasitized areas (e.g., gills and guts) and areas of important morphological characters (e.g., fins and lateral line area); from larvae and small specimens, it may be necessary to destructively sample a portion of the specimen, and consultation with taxonomists is advised so as to avoid critical morphological regions (e.g., the suction disk on clingfish).
4. Clean all tools prior to touching specimen using bleach solution or flame sterilization, etc. Scrape off scales from the area to be sampled and carefully dissect out a small portion of muscle—the amount of tissue to sample is dependent on the

size of the specimen and storage vessel (do not exceed a tissue:buffer ratio of 1:4 if possible) (see Note 3).

5. From EtOH-preserved and -stored specimens: DNA leached from the specimen can be extracted from the alcohol in the storage container (8). (After distillation, the used ethanol can be recycled without risk of contamination (7).) However, increased yields will be obtained via more substantial yet destructive tissue biopsy (as above in Subheading 3.1, step 2).

3.2. Tissue Storage

1. BioBanking: One of the significant contributions of the DNA barcoding enterprise is a repository of genetic materials. These are tied to voucher specimens in public collections and the identity and integrity of the specimens have been validated genetically. These materials can then serve as a starting point for subsequent molecular investigations. Therefore, it is important to maximize the utility of all collected materials.
2. Frozen storage: Freezing tissues (−20°C or lower) is the recommended method of preservation to maximize potential future uses of the material. Vapor-phase liquid nitrogen is ideal. Frequently this is not feasible, particularly in the field, so alternatives are presented.
3. Salt/DMSO buffer storage: Transportation of ethanol and other flammables has become an issue, and salt/DMSO buffer is an option in those cases. Place small tissue chunks in the buffer, taking care not to overwhelm the buffer with too much tissue—a good ratio of tissue:buffer is 1:4.
4. Ethanol (95%) storage: 70% ethanol should be avoided (see Note 1).

3.3. DNA Extraction: Autogen Prep 965

1. Prepare fresh lysis solution for every run by dissolving appropriate aliquots of proteinase-K provided in kit with each aliquot of Reagent M1. Standard concentration of proteinase-K in M1 lysis buffer for overnight digestion of animal tissue is 0.4 mg/ml.
2. For tissue lysis, place tissue into the appropriate well of a 96 deep-well plate (Costar #3960), and add 150 μl of Reagent M2 and 150 μl Reagent M1 containing the predissolved proteinase-K at the concentration of 0.4–1.0 mg/ml. Cover the plate with a silicone mat (Axygen) and one or more plexiglass plates cut to fit the block to minimize evaporation of buffer and prevent contamination between wells. The silicone mat and plexiglass plates are taped firmly to the block.
3. Incubate the samples overnight at 56°C with shaking.
4. Spin the plates briefly to remove condensed droplets from the lid. Load lysis plates on the AutoGenprep 965, with an equal number of output plates for DNA and tips following manufacturer's instructions. A maximum of four 96-well plates can be run simultaneously on the machine (see Note 4).

5. Load Reagents R3, R4, R5/R6/R7, R8, and R9 into the appropriate reservoirs following manufacturer's instructions.
6. The standard resuspension volume is 0.05 ml of buffer R9. If a large quantity of DNA is expected, such as when extracting vertebrate tissues or large amounts of other tissue, we change this volume to 0.1 ml.
7. Run the protocol on the instrument.
8. Upon completion, using a 96-well pipettor (if available), portion the DNA extracts into the desired amounts for the working and archival stocks for BioBanking.

3.4. DNA Extraction: Qiagen Biosprint Magnetic Bead Protocol (96-Well Plate Protocol)

1. We follow the Qiagen protocol for Biosprint extractions. Following are our deviations from the published protocol and our observations on particular steps.
2. Before beginning, check buffer ATL for white precipitate and take steps to resuspend it (see Note 6).
3. The MagAttract particles settle out of solution very quickly. Before adding this mix to the master mix, vortex at high speed for 3 min. Use immediately. Vortex again if much time has elapsed (>2–3 min).
4. Prepare Master mix of AL Buffer = 100 µl; isopropanol = 100 µl; MagAttract Suspension G = 15 µl. Prepare master mix 10% greater than that required for the total number of sample purifications to be performed.
5. Cut 5–25 mg of each tissue into small pieces and place in a 96-well S-Block. Add buffer ATL and proteinase-K.
6. Seal the plate following the same method as the Autogen lysis above (Subheading 3.3, step 2).
7. Place sealed plate in an incubator/shaker and digest overnight at 56°C (see Note 7).
8. Following lysis with ATL, briefly centrifuge the S-Block containing the samples to remove drops from underneath the lid.
9. Vortex the master mix containing Buffer AL, isopropanol, and MagAttract Suspension G (see Note 6) for at least 1 min. Add 215 µl of this master mix to each sample in the S-Block.
10. Place blocks on instrument and start and run protocol.
11. Upon completion, using a 96-well pipettor (if available), portion the DNA extracts into the desired amounts for the working and archival stocks for genetic repository (see Note 8).

3.5. Manual Extraction: Phenol:Chloroform Protocol, Following Ref. 20

1. It is typically easiest to carry out the extraction out in 1.7–2-ml centrifuge tubes.
2. For lysis, prepare fresh stock of lysis buffer (from Subheading 2.3, item 5) and add proteinase-K to 1 mg/ml.
3. Combine 1 ml lysis buffer with ~1 cm^2 of tissue.

4. Incubate with shaking overnight at 56°C.
5. Add no more than 700 µl of the lysed sample to the extraction tube. If a large amount of tissue is being extracted, it may be best to dilute the extraction with more extraction buffer, divide the sample into more than one tube, and extract each separately (see Note 5).
6. Phenol extraction: Add an equal volume of phenol to the lysed sample, and vortex vigorously to mix the phases. Spin in a microcentrifuge at top speed for 1–2 min to separate the phases. Remove aqueous phase (top layer) to new tube, being careful to avoid the phase interface.
7. Repeat the phenol extraction two more times.
8. Chloroform:isoamyl extraction: As in Subheading 3.5, step 6, use an equal volume of chloroform:isoamyl alcohol (instead of phenol) to remove any trace phenol. Repeat once more.
9. Precipitate the DNA with equal volume of isopropanol, and incubate at –20°C for 1 h or longer.
10. Spin sample with alcohol in microcentrifuge for 10 min at maximum speed to pellet DNA.
11. Before resuspension, set tube on counter at room temperature (RT), covered with a tissue, for 10 min or until all residual ethanol has evaporated.
12. Resuspend DNA pellet in a volume (usually, 50–200 µl) of TE (or a 1/10 dilution of TE) to achieve desired concentration. Vortex to mix.

3.6. PCR Methods: Amplification

1. Thaw and prepare reagents in proper concentration for the PCR reaction. Wait for each reagent to thaw completely and then mix thoroughly. Dilute primers to a 10 µM working stock for the PCR reaction. If starting with lyophilized primers, spin down before opening the tube and resuspend to 100 mM in molecular-grade water to form the stock solution. Dilute this stock 1:10 to make working stock.
2. Mix all reagents in volumes listed in the PCR recipe (Table 2) to form the master mix. Keep your reaction plate (or tubes) and master mix on ice. Vortex the master mix vigorously or pipette up and down to mix well (vortexing will cause liquid to be trapped on the cap of the tube, so follow with 15-s spin in a mini-centrifuge).
3. Aliquot 9 µl of master mix into each well of the 96-well plate (Genemate).
4. Aliquot 1 µl of each DNA template (undiluted) to each well. Change tips between samples.

Table 2
PCR reaction cocktail

PCR reagents	Each well (μl)	96-Well plate (μl)
ddH_2O	6.4	640
10 mM dNTPs	0.5	50
10× buffer	1	100
50 mM $MgCl_2$	0.4	40
10 μM primer F	0.3	30
10 μM primer R	0.3	30
Bioline Taq (5 U/μl)	0.1	10
Total	9	900

5. Add 1 μl nuclease-free water to well H12 to function as negative PCR control.
6. Place silicone plate mat over the top of the 96-well plate and secure it with a roller. Centrifuge the plate in a plate centrifuge or plate spinner for 10–15 s at approximately 3,950 rcf (= "centrifuge briefly").
7. Place the plate in a thermal cycler block and run PCR with the following cycling parameters: 95°C for 5 min, 35 cycles of 95°C for 30 s, 52°C for 30 s, and 72°C for 45 s, with a final extension at 72°C for 5 min and hold indefinitely at 10°C (see Notes 9–11).

3.7. PCR Methods: Visualization via Agarose Gel Electrophoresis

1. Cast a 1.5% agarose gel for each PCR reaction plate (e.g., add 0.75 g agarose to 50 ml 1× TBE buffer and boil in the microwave until the agarose is dissolved) (see Note 12).
2. Cool the solution, then add 1 μl ethidium bromide (10 mg/ml) per 50 ml of agarose solution, mix well, and pour immediately into the casting tray.
3. Let gel set for approximately 30 min or until firm. Remove combs from the gel and place it into an electrophoresis rig filled with 1× TBE buffer.
4. Add 2 μl 2× loading dye to each well of an empty 96-well plate.
5. Add 2 μl of each PCR product to the loading dye plate. Use new tips for each transfer.
6. Mix PCR product and dye by pipetting up and down, and then load 4 μl of the mixture to each well of the gel (see Note 12).

Table 3
Exosap purification of PCR product

Reagents	Each well (μl)	96-Well plate (μl)
ddH_2O	1.5	150
ExoSAP-IT	0.5	50
Total	2	200

7. Run gel at 100 V for approximately 12 min or until the bromophenol blue and xylene cyanol dyes in the loading dye are clearly separated.
8. At the end of the run, remove the gel from the rig and place on a UV transilluminator. Use a gel-imaging system to capture a digital image of the gel.

3.8. PCR Purifications: EXOSAP-IT

1. We use a fourfold dilution of the ExoSAP-IT mix. In a 1.7-ml microcentrifuge tube, mix nuclease-free water and ExoSAP-IT in the volumes listed in the Table 3. Keep the enzyme mix on ice or cold block at all times (see Note 13).
2. Vortex the diluted mix vigorously or pipette up and down to mix well.
3. Centrifuge the plate briefly to bring down all condensation from the sides of the wells and lid to the bottom.
4. Aliquot 2 μl of ExoSAP-IT mix to each well of the 96-well plate.
5. Place the silicone plate mat back on the top of the 96-well plate, press it down with a roller, and centrifuge briefly.
6. Place the plate in a thermal cycler block and run with the following parameters: 37°C for 30 min and 80°C for 20 min, and hold on 10°C.

3.9. Sanger Cycle Sequencing Protocol

1. Thaw and prepare reagents in proper concentration for the cycle sequencing reaction. Wait for each reagent to thaw completely and then mix them thoroughly. Keep the BigDye on ice and in the dark since the BigDye is both light and temperature sensitive.
2. Mix all four reagents in the volumes as listed in Table 4. Create two master mixes, one containing the forward primer and one containing the reverse (see Note 14).
3. Vortex the mixes vigorously or pipette up and down to mix. Centrifuge briefly.
4. Aliquot 9 μl of master mix into each well of the 96-well plate, keeping both plate and master mix on ice.

Table 4
Cycle sequencing reaction recipe

Reagents	Each well (μl)	96-Well plate (μl)
ddH_2O	25	625
5× SEQ buffer	1.75	175
BigDye	0.5	50
10 μM primer F or R	0.5	50
Total	9	900

5. Aliquot 1 μl of purified PCR product into each well of the 96-well plate. Change tips after each sample.
6. Place a silicone plate mat on the 96-well plate, press it down with a roller, and centrifuge briefly.
7. Place the plates in a thermal cycler and run the cycle sequencing program with the following parameters: 30 cycles of 95°C for 30 s, 50°C for 30 s, and 60°C for 4 min and hold at 10°C.
8. After the program has finished, store the plates at 4°C in a dark refrigerator until purification (see Note 14).

3.10. Sephadex Purification of Sanger Sequence Reactions

1. Measure dry Sephadex G50 using the multiscreen black column loader (predrilled uniform holes which measure and deliver the correct amount of sephadex) into a Multiscreen® HTS filter plate for each 96-well cycle sequencing reaction plate to be cleaned.
2. Add 300 μl molecular-grade water to the wells and allow to sit at room temperature for at least 2 h in order to completely hydrate the sephadex matrix.
3. Place the filter plate on top of a 96-well PCR plate (the "catch" plate) and tape together with laboratory tape.
4. Centrifuge at 750 × *g* force for 5 min to drain the excess water from the wells. Discard the water. The catch plate can be used again, but ONLY for this purpose, not PCR.
5. Add the entire volume of the sequencing reaction to the center of the Sephadex columns (not along the side walls) taking care not to touch the column surface or destroy the integrity of the column.
6. Attach a *new* 96-well PCR plate to the bottom of each Sephadex plate and secure with lab tape. Make sure that the orientation of the plates is the same: the A1 well of the Sephadex plate is over the A1 well of the catch plate.

7. Centrifuge at 750 × *g* force for 5 min to elute the cleaned sequencing product into the catch plate.
8. Dry the cleaned sequencing products on a heat block at 90°C for 10–15 min or in a Sorvall Speedvac. Cover the plate with Septa Seal mat and store at –20°C (see Note 15).
9. To prep for running on the genetic analyzer, add 10 μl Hi-Di™ Formamide to each well (under a fume hood).
10. Denature the DNA by heating the plate at 97°C for 3 min.
11. Cool the plate at 4°C for 3 min.
12. Load the plates into the Genetic Analyzer.

3.11. Genetic Analyzer Methods

1. ABI 3130XL: Running a 36-cm capillary array run using the ABI Template Protocol "RapidSeq36_POP7" with a run time of 2,280 s.
2. ABI 3730XL: Running a 50-cm capillary array run using the ABI Template Protocol "LongSeq50_POP7" with a run time of 4,000 s (see Note 16).

3.12. Data Processing and Quality Control Methods

1. Production of the final DNA barcode sequence from the raw sequencer output (the "traces") involves several steps (forward and reverse traces for one set of 96 specimens are processed together).
2. The method we outline here uses Sequencher vers. 4.10.1 (Gene Codes, Corp). Alternatively, use of the Geneious program (BioMatters) and the BioCode plugin is discussed elsewhere in this volume.
3. Trace trimming—trimming is based on the phred quality scores of each base call (see Note 17). Trimming criteria, as implemented in Sequencher, are as follows: trim from the 5′ and 3′ ends until the first (or last) 20 bases contain fewer than 3 bases with a phred score <20, and from both ends until the first (or last) 10 bases contain <3 ambiguous bases.
4. Filtering traces—after trimming, all traces <450 bp in length are discarded, as are those where <90% of the bases have an average phred score of 20 or higher (see Note 18).
5. Contig building—using the "Assemble by Name" option in Sequencher (if reference sequence is used, then "Assemble by Name to Reference"; see Note 19 for our degenerate reference sequence), forward and reverse trimmed traces are assembled into contigs.
6. If the two reads are less than 97% identical, the sequences are set aside for manual interpretation of cause (see Note 20). Such dissimilarity may be due to low-quality traces (if both reads are on the lower end of our quality spectrum, a good contig may not be possible), laboratory error, sequencer error or contamination.

Manual base calls in the consensus sequence resulting from these contig assemblies are based on "confidence" (phred scores).

7. If there is a discrepancy in a base call between the forward and reverse reads, the base call in the consensus sequence will reflect the base with the highest confidence of the two (under the CONTIG menu, choose "Consensus by Confidence) (see Note 21).
8. Inspection—contig assemblies are manually inspected if any ambiguities or gaps exist or if there are more than three disagreements between the two reads (this information is available using Sequencher's "Get Info" option). Base calls are double checked by eye and left as ambiguities when neither base call has high confidence (see Note 22).
9. Finalizing the consensus sequence—all reads are assembled together with a reference sequence into one large assembly (usually using the lowest similarity requirement—60%). The reference sequence is used as a length gauge and primer sequences are trimmed from all forward and reverse reads at this point, if not done in prior step.
10. The assembly is examined for gaps, which at this point usually stem from a base call error (e.g., one base mistakenly interpreted by the software as two), and these mistakes are rectified. The large assembly is then "dissolved" and each specimen's pair of traces reassembled as in step 3. The consensus sequence is then exported from each trace contig assembly as a concatenated fasta-formatted text file.
11. Contamination check—if, in the previous step, a read does not assemble, then it is compared to the NCBI nucleotide database using BLAST (see Note 23).

4. Notes

1. Denatured ethanol is not conducive to molecular work and should be avoided.
2. If "painting" fins with formalin to better photograph a specimen's natural colors, the biopsy should be taken from right side first or from an area not exposed to any formalin, as this can hinder DNA extraction and/or amplification.
3. When sampling specimens in the field, cutting the larger tissue sample into smaller pieces suitable for DNA extraction saves time back in the lab. Upon returning, one then only needs to open the tube and pick out one or two small minced pieces to place into the tube/plate for cell lysis and DNA extraction. Not doing this means the larger (pencil eraser size) piece of

tissue must be removed from the tube and subsampled at a later time, requiring more tools, sterilization, and, most importantly, more time. Partially mincing the tissue also aids penetration of the stabilizing buffer/solution.

4. There is a minimum of 2 plates and 24 samples (12 samples of Row A in each plate) required to balance the built-in centrifuge of the Autogen 965.
5. It is difficult to extract volumes smaller than 100 μl. A sample can be concentrated by resuspending in a smaller volume after precipitation.
6. Check that Buffer ATL does not contain a white precipitate. If necessary, incubate for 30 min at 37°C with occasional shaking to dissolve precipitate. MagAttract magnetic particles copurify RNA and DNA if both are present in the sample. If RNA-free DNA is required, add RNase A to the sample before starting the procedure. The concentration of RNase should be 2 mg/ml (add 2 μl of 100 mg/ml RNase A solution to each 100 μl of sample).
7. Place a weight on top of the caps during incubation. Mix occasionally during incubation to disperse the sample or place on a rocking platform. Lysis time varies depending on the type, age, and amount of tissue being processed. Lysis is usually complete in 1–3 h, but optimal results will be achieved after overnight lysis. After incubation, the lysate may appear viscous, but should not be gelatinous as it may clog the DNeasy 96 membrane: dilution may be required.
8. Elution with volumes less than 200 μl increases the final DNA concentration but may reduce the overall DNA yield. For samples containing less than 1 μg DNA, elution in 50 μl Buffer AE is recommended. For maximum DNA yield, repeat step 16 with another 200 μl Buffer AE. A second elution with 200 μl Buffer AE will increase the total DNA yield by up to 25%. However, due to the increased volume, the DNA concentration is reduced. If a higher DNA concentration is desired, the second elution step can be performed using the 200 μl eluate from the first elution. This will increase the yield by up to 15%.
9. Tailed primers (e.g., FISHCO1LBCm13F and FISHCO1-HBCm13R) can be used in the PCR reaction for easier downstream processing. In this case, the sequencing primers will be the m13F and m13R primers.
10. If the PCR amplification is not successful, we use the standard animal 16S primers (16Sar and 16Sbr) (21) to check the quality of the DNA. The PCR recipe and cycling parameters are the same as for CO1, but with an annealing of 48°C instead of 52°C.
11. Silicone plate mats can be cleaned by soaking for a minimum of 10 min in a 10% bleach solution followed by rinsing with

distilled water. After drying, UV irradiation is recommended prior to reuse.

12. Agarose gels can be stored for 2–3 days prior to use. To store, wrap gels in plastic wrap and keep at 4°C. Gels can also be made up in large volumes and stored in aliquots of a volume sufficient for pouring one gel (in 50-ml Falcon tubes, which can be reused). At the time of use, the solid gel is put into a beaker and microwaved until liquid.
13. The original strength Exosap-IT can be diluted as much as tenfold and still be effective; this will cut down costs even further. Incubation time at 37°C must be increased to accommodate the lower quantity of enzyme. Diluted ExoSAP-IT mixes cannot be refrozen and has a short shelf life at 4°C. We discard any excess diluted ExoSAP-IT immediately.
14. Keep sequencing reactions in a dark place at all times to avoid degradation of the light-sensitive sequencing product. If tailed primers FISHCO1LBCm13F and FISHCO1HBCm13R were used in the PCR reaction, the sequencing primers can be the same as the amplification primers or, alternatively, the m13 sequencing primers m13F and m13R.
15. Multiscreen® HTS filter plates (as well as the plastic catch plates used for Sephadex) can be reused by filling the filter plate with 300 µl distilled water and spinning for 5 min at 700 × *g* force. Repeat this procedure once more and invert to dry. Dried cleaned sequencing reactions can be kept for months at –20°C or can be shipped. While other methods than Sephadex cleanup can be used and are sometimes less expensive (e.g., ethanol precipitations), improper execution by inexperienced users can lead to premature degradation of the capillary arrays on the genetic analyzer. In the hands of most users, the Sephadex method consistently produces high-quality cleaned sequencing products.
16. The newest polymer from Applied Biosystems, POP7, is used for all sequencing runs, and the running buffer is changed 3×/week. The number of samples in the 3730XL queue can be increased by combining four sets of 96 samples into a 384-well plate. This can maximize the number of samples run over a weekend. Once run, the sequencing reactions are kept at 4°C for a week. They can be re-run on the sequencer with no significant loss of signal during this period.
17. Phred scores are essentially a measure of the accuracy of a base call made by the sequencing software. According to many sources, a phred score of 20 indicates a 99% probability of an accurate base call. We consider base calls with a phred score of <20 as "low" quality and "high"-quality calls have a phred score >40 (users with different parameter definitions can generate incongruent results).

18. Occasionally, sequences with an initial overall quality <80% (see Subheading 3.7, step 2) can be improved via visual inspection and manual trimming. We only do this when it is clear that the trimming parameters failed to remove additional poor-quality bases at the end of the sequence, thus lowering the overall quality. It is sometimes possible to cut out the poor-quality bases and raise the overall sequence quality while maintaining a length >450 bp. In Sequencher, candidates for this treatment may be recognized as post-trim sequences of adequate (or longer) length with 70–80% overall quality.
19. Our degenerate fish reference sequence is 5′ → 3′ NNTTTAT NTAGTATTTGGTGCCTGAGCCGGAATAGTAGGCACA GCCCTAAGCYTAYTAATTCGAGCTGAACTAAGCCAAC CTGGCGCCCTNNTNGGNGACGACCAAATTTATAATG TAATCGTAACTGCCCACGCCTTTGTAATAATTTTCTT TATAGTAATACCAATTATGATTGGAGGCTTTGGAAAC TGAYTAATCCCCCTAATGATTGGGGCCCCCGACATGG CCYTCCCYCGAATAAACAACATAAGCTTTTGNNTNNT NCCNCCNTCNNTCNTNNTNNTNNTNGCATCCTCTG GNNTNGAAGCCGGGGCCGGAACAGGATGAACAGTT TANCCNCCNNTAGCNGGAAACYTAGCCCACGCAGGA GCCTCTGTAGACCTAACAATTTTCTCCYTTCATCTAG CAGGAATNTCCTCAATNNTNGGNGCAATTAACTTTA TTACAACAATYYTNAANATGAAACCNCCNGCNATNTC NNNNTACCAAACACCCYTATTTGTTTGAGCNNNNYT AATTACAGCCGTNNTNNTNNTNNTNTNNYTCCNNG TCYTTGCTGCTGGCATTACAATGYTTYTCACAGACCG AAACYTAAACACAACCTTCTTTGACCCTGCAGGAGGA GGAGACCCCATTYTGTACCAACACYTAYTC.
20. If a contig has a few disagreements between the sequences but no ambiguities or gaps, each base in the sequence has a "solid" call based on relative confidence. Contigs with >3 disagreements and no ambiguities are visually inspected to ascertain the cause(s) of the disagreement.
21. To use the "Assemble by Name" option in Sequencher which is really the only way to create a large number of assemblies without a tedious amount of file renaming, it is important to plan ahead. Trace files need to come from the sequencer with names that can be readily combined; therefore, files from a run of 96 using multiple primers or plates need to be named in a similar fashion so that the entire batch can be treated at once. Standardization is the key. Our fish trace files are labeled field-number_Genus_species_barcode_F (only part of the generic and specific names are used; e.g., BLZ8013_Scaru_iseri_89893427_F, a forward trace file). Sequencher breaks up a name as directed—underscores are most common (the above

example would break into five names). We often use regular expressions to define the parameters to be used for the splitting so that we can get the names we want on our final fasta files (e.g., (.*)_(.*) will take off only the F or R from the above example and the remainder will be the consensus sequence fasta file defline).

22. At the point of visual inspection, there are generally no changes to be made to the contig. All of the previous stringent filtering, trimming, etc. leave very few miscalls. The most common errors remaining are multiple calls—2 real As are called as 3, 3 real Gs are called as 2. Errors such as these can cause a gap to be inserted in the consensus sequence. Miscalls such as these can slip through all of our QC steps and are quite pernicious. Errors, where one extra base is called, can be caught by assembling all sequences in a project together with a reference sequence. But errors, where too few bases are called, are more vexing. There is no foolproof method for identifying them. When they occur very near the 3′ end of a sequence, if not caught, they can lead to artificial increases in distances from other sequences or stop codons.
23. In general, sequences from almost all fish species will readily assemble to one another and if a sequence does not, then it is likely a contaminant. Contaminants in our fish sequences are generally bacterial or fungal in origin. BLASTing against GenBank is the best way to verify this, as these contaminating sequences are often not a match to any known COI sequence.

Acknowledgments

Our Fish DNA Barcoding Team would like to thank the NMNH Fish scientists and taxonomists for taking the time to educate us on fish biology and biodiversity and make this work thoroughly enjoying. We also thank David Erickson for his tireless efforts to produce this important volume.

References

1. Eschmeyer WN, Fricke R, Fong JD, Polack DA (2010) Marine fish diversity: history of knowledge and discovery (Pisces). Zootaxa 2525: 19–50
2. Handy SM, Deeds JR, Ivanova NV et al (2010) A single laboratory validated method for the generation of DNA barcodes for the identification of fish for regulatory compliance. J AOAC Int 94:1–10
3. Yancy HF et al (2008) A protocol for validation of DNA-barcoding for the species identification of fish for FDA regulatory compliance. FDA Lab Inf Bull LIB No 4420
4. Pilgrim EM, Jackson SM, Swenson S, Turcsanyi I, Friedman E, Weigt LA, Bagley MJ (2010) Incorporation of DNA barcoding into a large-scale biomonitoring program: opportunities and pitfalls. J NABS 30:217–231

5. Gemeinholzer B, Rey I, Weising K, Grundmann M, Muellner AN et al (2010) Organizing specimen and tissue preservation in the field for subsequent molecular analyses, chap. 7. In: Eymann J, Degreef J, Hauser C, Monje JC, Samyn Y, Vanden Spiegel D (eds) Manual on field recording techniques and protocols for all taxa biodiversity inventories. ABCTaxa (Belgium) 8:653
6. Leponce M, Meyer C, Hauser CL, Bouchet P, Delabie J, Weigt L, Basset Y (2010) Challenges and solutions for planning and implementing large-scale biotic inventories, chap. 3. In: Eymann J, Degreef J, Hauser C, Monje JC, Samyn Y, VandenSpiegel D (eds) Manual on field recording techniques and protocols for all taxa biodiversity inventories. ABCTaxa (Belgium) 8:653
7. Keel WG, Moser W, Giaccai J, Ormos A, Tanner J, Weigt LA (2011) Alcohol recycling at the Smithsonian Institution, National Museum of Natural History (NMNH). Collect Forum 25(1):10–21
8. Shokralla S, Singer GAC, Hajibabaei M (2010) Direct PCR amplification and sequencing of specimens' DNA from preservative ethanol. Biotechniques 48:305–306
9. Venegas RP, Hueter R, Gonzalez J, Tyminski J, Remolina JFG et al (2011) An unprecedented aggregation of whale sharks, *Rhincodon typus*, in Mexican coastal waters. PLoS One 6(4): e18994.
10. Baldwin CC, Mounts JH, Smith DG, Weigt LA (2009) Genetic identification and color descriptions of early life-history stages of Belizean *Phaeoptyx* and *Astrapogon* (Teleostei: Apogonidae) with comments on identification of adult *Phaeoptyx*. Zootaxa 2008:1–22
11. Baldwin CC, Weigt LA, Smith DG, Mounts JH (2009) Reconciling genetic lineages with species in Western Atlantic *Coryphopterus* (Teleostei: Gobiidae). Smithsonian contribution. Mar Sci 38:111–138
12. Baldwin CC, Castillo CI, Weigt LA, Victor BC (2011) Seven new species within western Atlantic *Starksia atlantica*, *S. lepicoelia*, and *S. sluiteri* (Teleostei: Labrisomidae), with comments on congruence of DNA barcodes and species. Zookeys 79:21–72
13. Tornabene L, Baldwin CC, Weigt LA, Pezold F (2010) Exploring the diversity of western Atlantic *Bathygobius* (Teleostei: Gobiidae) with cytochrome c oxidase-I, with descriptions of two new species. Aqua 16:141–170
14. Wong EH-K, Shivji MS, Hanner RH (2009) Identifying sharks with DNA barcodes: assessing the utility of a nucleotide diagnostic approach. Mol Ecol Res 9:243–256
15. Hajibabaei M, Smith MA, Janzen DH, Rodriguez JJ, Whitfield JB, Hebert PDN (2006) A minimalist barcode can identify a specimen whose DNA is degraded. Mol Ecol Notes 6:959–964
16. Kumar R, Singh PJ, Nagpure NS, Kushwaha B, Srivastava SK, Lakra WS (2007) A non-invasive technique for rapid extraction of DNA from fish scales. Ind J Exp Biol 45:992–997
17. Lucentini L, Caporali S, Palomba A, Lancioni H, Panara F (2006) A comparison of conservative DNA extraction methods from fins and scales of freshwater fish: a useful tool for conservation genetics. Cons Genet 7:1009–1012
18. Ivanova NV, Dewaard JD, Hebert PDN (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. Mol Ecol Notes 6:998–1002
19. Sambrook J, Fritsch EF, Maniatis T (1989) Purification of nucleic acids. In: Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, pE.3–4
20. Palumbi SR (1996) Nucleic acids II: the polymerase chain reaction. In: Hillis DM, Moritz C, Mable BK (eds) Molecular systematics. Sinauer, Sunderland, MA
21. Seutin G, White BN, Boag PT (1990) Preservation of avian blood and tissue samples for DNA analysis. Can J Zool 69:82–90
22. Ivanova NV, Zemlak TS, Hanner RH, Hebert PDN (2007) Universal primer cocktails for fish DNA barcoding. Mol Ecol Notes 7:544–548
23. Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann Entomol Soc Am 87:651–701