

Environmental Genomics

Edited by

C. Cristofre Martin



Assembling DNA Barcodes

Analytical Protocols

**Jeremy R. deWaard, Natalia V. Ivanova, Mehrdad Hajibabaei,
and Paul D. N. Hebert**

Summary

The Barcode of Life initiative represents an ambitious effort to develop an identification system for eukaryotic life based on the analysis of sequence diversity in short, standardized gene regions. Work is furthest advanced for members of the animal kingdom. In this case, a target gene region has been selected (cytochrome *c* oxidase I) and pilot studies have validated its effectiveness in species discovery and identification. Based on these positive results, there is now a growing effort to both gather barcode records on a large-scale for members of this kingdom and to identify target barcode regions for the other kingdoms of eukaryotes. In this chapter, we detail the protocols involved in the assembly of DNA barcode records for members of the animal kingdom, but many of these approaches are of more general application.

Key Words: Biodiversity; COI; *cytochrome c oxidase I*; DNA barcoding; DNA sequencing; mitochondria; species identification; taxonomy.

1. Introduction

The DNA barcoding movement seeks to advance biodiversity science through the development of DNA-based systems that aid species identification and discovery (*1–4*). In particular, it aims to build these systems in a parsimonious fashion, by basing them, whenever possible, on sequence diversity in a short, standardized gene region. DNA barcoding is a young endeavor whose activation traces to a 2003 publication (*1*) that revealed the likelihood of developing an effective system for species identifications in the animal kingdom based on

From: *Methods in Molecular Biology*, vol. 410: *Environmental Genomics*
Edited by: C. Cristofre Martin © Humana Press, Totowa, NJ

sequence variation in a 648 base pair (bp) region of the *cytochrome c oxidase I (COI)* gene. A number of studies have now validated this approach in varied taxonomic groups and geographic settings. Studies on North American birds have progressed to the point that barcode records are available for more than 95% of the species that breed on this continent (5,6). This work has revealed that DNA barcoding is effective; it provides species-level identifications for 94% of the 693 species analyzed and the few cases of compromised resolution regularly involve species pairs that are known to hybridize. The studies also affirm the value of DNA barcoding in species discovery; flagging 15 overlooked species of birds despite the intensity of prior taxonomic work on this continent. Fishes have also attracted substantial interest and barcode records are now available for more than 1000 species, with the most detailed study revealing 100% success in the identification of 250 Australian fish species (7). Work has also been carried out on a number of invertebrate lineages. Studies on molluscs have largely been restricted to members of a single family in which 94% success in species identification was reported (8). However, arthropods have been examined more intensively with studies on crustaceans (9), spiders (10), collembolans (11), and varied insect lineages including ants (12), mayflies (13), mosquitoes (14), and tachinid flies (15). Work on Lepidoptera has revealed the effectiveness of barcoding in both the discovery and identification of species in hyperdiverse tropical biotas (16,17). As well, wide-ranging studies on more than 1000 North American species of Lepidoptera revealed that regional variation in barcode sequences pose no difficulty for the development of an effective DNA-based identification system (18).

The ultimate goal of the DNA barcode movement is the development of comprehensive barcode libraries for all lineages of eukaryotes. While past work has focused on animals, investigations are underway to develop protocols for protists, plants, and fungi. There is evidence that the region of *COI* targeted in animals may also be effective as a basis for identification systems in algae (19) and fungi (20). In plants, a different genic target will be required, but exploratory efforts are underway (21,22). While protocol development is the primary area of endeavor in other kingdoms of eukaryotes, work on animals is scaling up through major barcode campaigns (23). Some of these campaigns have a taxonomic focus, such as the efforts to gather barcode records for all species of birds and fishes. In other cases, barcode campaigns seek to build a comprehensive barcode library for the biota in a particular region. For example, efforts are underway to assemble barcode records for 10,000 Canadian animal species (10% of fauna) within 5 years. The progress toward the large-scale activation of barcoding is now coordinated by the Consortium for the Barcode of Life, headquartered at the Smithsonian's National Museum of Natural History in Washington, but representing an alliance of more than 100 international organizations with interests in biodiversity science.

The growing intensity of barcode research has created the need for simple, inexpensive protocols for barcode analysis (24). This chapter describes the protocols employed in a DNA Barcoding Centre that processes approx 100K specimens a year. However, because the primary considerations for protocol adoption by this facility were simplicity, cost minimization, and speed, these analytical approaches represent a generic solution to barcode acquisition. In this chapter, we consider the eight key steps in the transition from the collection of an organism to the injection of its sequence record into either the Barcode of Life Data System or one of the major genomics repositories.

2. Materials

2.1. Specimens and Tissue Handling

1. 96-Well format 2-D barcoded storage tubes (TrakMates, Matrix Technologies).
2. 99.9% Ethyl alcohol (Commercial Alcohols). Store in a flammable liquids cabinet.
3. Forceps and/or scalpel (FineScience Tools).

2.1.1. Genomic DNA Extraction/Purification—Fresh or Frozen Tissue

1. Dry Release extraction buffer: 5–6 g of Chelex-100 (Bio-Rad Laboratories, Hercules, CA), 10 mL of 1% sodium azide (Sigma-Aldrich, St. Louis, MO), 1 mL of Tris-HCl, pH 8.3, ultrapure H₂O to 100 mL. Store at 4°C in 10-mL aliquots.
2. 20 mg/mL of proteinase K (Invitrogen, Carlsbad, CA): 100 mg of proteinase K, 5 mL of ultrapure H₂O. Store in 0.1- to 1-mL aliquots at –20°C.
3. DryRelease working solution: 10 µL of proteinase K, 100 µL of DryRelease extraction buffer.
4. Microplate (Eppendorf, New York, NY).
5. Cap strips (ABgene, Rochester, NY).
6. Thermocycler (Mastercycler EP Gradient, Eppendorf).

2.1.2. Genomic DNA Extraction/Purification—Archival Specimens

1. NucleoSpin 96 Tissue Kit (Macherey-Nagel).
2. 99.9% Ethyl alcohol (Commercial Alcohols). Store in a flammable liquids cabinet.
3. Microplate (Eppendorf).
4. Cap strips (ABgene).
5. Matrix ImpactII P1250 pipettor (Matrix Technologies).
6. Centrifuge with deep-well plate rotor (25R, Beckman Coulter).
7. Incubator (Fisher Scientific).

2.2. Polymerase Chain Reaction (PCR) Amplification of the Barcode Region

1. 10% Trehalose: 5 g of D-(+)-trehalose dehydrate (Sigma-Aldrich), ultrapure H₂O to 50 mL. Store at –20°C in 1- to 2-mL aliquots.

2. 10× PCR Buffer (New England Biolabs). Store at -20°C .
3. 50 mM MgCl_2 : 2 mL of 1 M MgCl_2 (Sigma-Aldrich), 38 mL of ultrapure H_2O . Store at -20°C in 1-mL aliquots.
4. 10 mM dNTP mix (New England Biolabs). Store at -20°C in 100- μL aliquots.
5. 100 μM primer stock: Dissolve desiccated primer (Invitrogen) in the amount of ultrapure H_2O indicated by the manufacturer to produce a final solution of 100 μM (i.e., add number of nmol \times 10 μL of ultrapure H_2O). Store at -20°C .
6. 10 μM primer working solution: 20 μL of 100 μM primer stock, 180 μL of ultrapure H_2O . Store at -20°C .
7. *Taq* polymerase (New England Biolabs). Store at -20°C in 50- μL aliquots.
8. Microplate (Eppendorf).
9. Cap strips (ABgene).
10. Thermocycler (Mastercycler EP Gradient, Eppendorf).

2.3. PCR Product Check

1. Mother E-base (Invitrogen).
2. 2% E-gel 96 gels (Invitrogen). Store at room temperature prior to opening and at 4°C between reuse. Agarose is encased in glass but does contain the mutagen ethidium bromide; waste should be disposed of according to local regulations.
3. Gel documentation system (AlphaImager 3400, Alpha Innotech Corp.).

2.4. Sequencing Setup

1. BigDye v.3.1. Cycle Sequencing Kit (Applied Biosystems). Store at -20°C .
2. 5× Sequencing buffer (Applied Biosystems). Store at 4°C .
3. 10% Trehalose: 5 g of D-(+)-trehalose dehydrate (Sigma-Aldrich), ultrapure H_2O to 50 mL. Store at -20°C frozen in 1- to 2-mL aliquots.
4. 100 μM primer stock: Dissolve desiccated primer (Invitrogen) in the amount of ultrapure H_2O indicated by the manufacturer to produce a final solution of 100 μM (i.e., add number of nmol \times 10 μL of ultrapure H_2O). Store at -20°C .
5. 10 μM primer working solution: 20 μL of 100 μM primer stock, 180 μL of ultrapure H_2O . Store at -20°C .
6. Microplate (Eppendorf).
7. Cap strips (ABgene).
8. Thermocycler (Eppendorf).

2.5. Sequencing Reaction Cleanup

1. Sephadex G-50 (Sigma-Aldrich). Skin exposure or inhalation may be harmful; wearing a face mask and gloves when handling is recommended.
2. AcroPrep 96 0.45 μM GHP filter plates (Pall Corp.).
3. Formamide (Applied Biosystems). Store at -20°C . This reagent is toxic; exposure should be minimized and waste should be disposed of according to local regulations.

4. Column loader (Millipore).
5. Centrifuge alignment frame (Millipore).
6. 96-Well reaction plates (Applied Biosystems).

2.6. Sequence Analysis

1. 96-Well septa (Applied Biosystems).
2. 96-Well plate bases (Applied Biosystems).
3. 96-Well plate retainers (Applied Biosystems).
4. 3730 Buffer (10×) with EDTA (Applied Biosystems). Store at 4°C.
5. POP-7 polymer (Applied Biosystems). Stable at room temperature for 7–10 days. This reagent is toxic; exposure should be minimized and waste should be disposed of according to local regulations.
6. 48-Capillary (50 cm) array (Applied Biosystems).
7. 3730 DNA Analyzer capillary sequencer (Applied Biosystems).

2.7. Sequence Editing/Alignment

1. Sequence editing software: Sequencher (Gene Codes, Ann Arbor, MI), SeqScape (Applied Biosystems), or Lasergene (DNASTAR, Madison, WI).

3. Methods

3.1. Specimens and Tissue Handling

Barcode analysis on most specimens is straightforward, but is highly dependent on the initial condition of the DNA. For this reason, care should be taken to ensure that specimens are killed in a DNA-friendly fashion, analysis should follow collection as quickly as possible (*see Note 1*), and precautions should be made to prevent contamination with DNA or other specimens (*see Note 2*). There are six key steps in ensuring the proper handling and documentation of each specimen:

1. Arrange specimens in batches of 94 (*see Note 3*). Laser-print small labels of unique specimen accession numbers for each of the 94 specimens. If a voucher specimen will be retained, affix a label to it or to the specimen container with this accession number.
2. Clean work surface with ethyl alcohol or a detergent for the removal of DNA and DNase contamination.
3. For each specimen, use acid- or flame-sterilized forceps and/or scalpel to remove a small tissue sample. Place tissue in an individual storage tube of the TrakMates box along with the corresponding specimen accession label.
4. Fill vials with ethyl alcohol for storage at room temperature. For dried tissue (e.g., insect legs) or tissue to be stored below freezing, the addition of ethanol can be omitted.

5. Record specimen accession numbers and associated data in electronic spreadsheet (see www.barcodeoflife.org for spreadsheet and documentation).
6. Photograph each specimen.

3.2. Genomic DNA Extraction/Purification

Numerous options are available for the extraction and/or purification of genomic DNA (see **Note 4**). Several of these methods have been thoroughly tested for their efficacy in high-volume animal DNA barcoding (**24**); the two methods that demonstrated superior performance for fresh/frozen tissue and for archival tissue, respectively, are outlined in the following two subheadings.

3.2.1. Genomic DNA Extraction/Purification—Fresh or Frozen Tissue

The Chelex-based DryRelease method (**24**) (see **Note 5**) is a quick and cost-effective method of DNA isolation, particularly useful for fresh or frozen material.

1. Aliquot 30–110 μL of DryRelease working solution (see **Note 6**) into each well of a microplate using a multi-channel pipet, wide-bore tips, and a reservoir. Mix solution before and while aliquoting to ensure that the resin is equally dispersed between wells. Cover each row with cap strips.
2. Put a tiny amount of tissue (e.g., 1–2 mm of insect leg or 1–2 mm^3 of ethanol-preserved tissue) into each well of the plate. To prevent cross-contamination, work with one row at a time. Close the lids and prevent shaking to ensure that the tissue fragment remains in the solution.
3. Incubate for 12–24 h at 55°C.
4. Centrifuge the plate at 1000g for 5 min. Incubate samples in a thermocycler at 95°C for 20 min to denature the proteinase K.
5. Store extractions at –20°C.
6. Before PCR set up, centrifuge the plate at 1000g for 5 min.
7. Use 1–2 μL of DNA sample for PCR. Ensure that Chelex resins are not transferred into the PCR reaction.

3.2.2. Genomic DNA Extraction/Purification—Archival Specimens

The NucleoSpin 96 Tissue Kit (Machery-Nagel) is a silica membrane-based method that performs well with both fresh and archival (older than 5 years) tissue (**24**) (see **Note 7**). The eluted DNA is highly purified, making this method ideal for isolating DNA for long-term storage.

1. Add a small amount of tissue (e.g., 2–4 mm of insect leg or 1–3 mm^3 of ethanol-preserved tissue) (see **Fig. 1**) to each well of the round-well block supplied with the kit. Maceration of the tissue is optional, but dividing the tissue into smaller pieces improves the final yield of DNA.

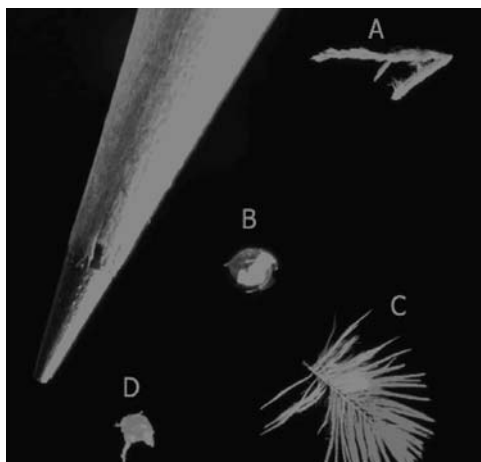


Fig. 1. Typical specimen sizes used for DNA extraction as compared to the head of a pencil. (A) Moth leg, (B) small crustacean (*Daphnia*), (C) bird feather, and (D) muscle tissue. (Reproduced from 24.)

2. Prepare a lysis working solution by combining 18 mL of buffer T1 with 2.5 mL of proteinase K in a reservoir. Using a multichannel pipet, transfer 200 μ L of the working solution into each well of the round-well block (see Note 8).
3. Seal wells with the cap strips provided and shake vigorously for 10–15 s to mix. Centrifuge at 1500g for 15 s to collect samples at the bottom of the wells.
4. Incubate at 56°C for a minimum of 6 h (ideally 18–24 h) to allow digestion. Tape down cap strips to prevent them from popping off.
5. After digestion, centrifuge at 1500g for 15 s to remove any condensate from the cap strips.
6. Premix ethanol and buffer BQ1: mix 20 mL of ethanol with 20 mL of buffer BQ1 in a reservoir. Using a multichannel pipet, transfer 400 μ L of the mixture into each well of the round-well block. Seal wells with cap strips. Shake vigorously for 10–15 s and centrifuge at 1500g for 10 s to remove any sample from the cap strips.
7. Remove cap strips and transfer lysate (about 600 μ L) from the wells of the round-well block into the wells of the tissue binding plate placed on top of a square-well block. Seal plate with a self-adhering PE foil supplied with the kit.
8. Centrifuge at 5600g for 10 min to bind DNA to the silica membrane.
9. Perform the first wash step: Add 500 μ L of buffer BW to each well of the tissue binding plate using a reservoir and multichannel pipette. Use a new self-adhering PE foil to seal the plate and centrifuge at 5600g for 2 min.
10. To accommodate the volume of flow through, replace the current square-well block with a new square-well block, placing it underneath the tissue binding plate.

11. Perform the second wash step: Add 700 μL of buffer B5 to each well of the tissue binding plate using a reservoir and multichannel pipet. Use a new self-adhering PE foil to seal the plate, then centrifuge at 5600g for 4 min.
12. Remove the self-adhering PE foil and place the tissue binding plate on a sterile microplate. Incubate at 70°C for 10 min to evaporate residual ethanol.
13. Dispense 30–100 μL ddH₂O directly onto the membrane of each well of the tissue binding plate (*see Note 9*). Incubate at room temperature for 1 min and seal plate.
14. Place both the tissue binding plate and the microplate on an open rack of MN tube strips (*see Note 10*). Centrifuge at 5600g for 2 min. Rotate the plate/microplate/rack 180° and repeat the centrifugation. Carefully remove the tissue binding plate and rack. Seal the microplate with cap strips.
15. Keep DNA at 4°C for temporary storage or at –20°C for long-term storage. Use 1–2 μL of the DNA sample for PCR amplification.

3.3. PCR Amplification of the Barcode Region

Barcode markers were chosen in part for their ease of isolation, satisfying such criteria as high copy number and the presence of conserved flanking regions for primers. PCR amplification of these regions is therefore routine with compliant samples and well-designed primers (*see Note 11*). For recalcitrant cases, success is often achieved by targeting smaller fragments or making small modifications to the primer sequence.

1. Defrost your reagents and place in a cold block.
2. Mix reagents in a 1.5-mL tube following the recipe given in **Table 1** (*see Notes 12–14*). Additives (*see Note 15*) or alternative enzymes (*see Note 16*) may increase the success, yield, and accuracy of the PCR. A list of primers to amplify the *COI* barcode region for a variety of taxonomic groups is given in **Table 2**.
3. Vortex-mix gently and aliquot 10.5 μL of the PCR mix into each well of the microplate. Time can be saved by aliquoting 1/8 (~138 μL) of the total mix into an eight-tube PCR strip and dispensing to the microplate with a multichannel pipet.
4. Add 0.5–2 μL of DNA extract (*see Notes 17 and 18*) to each well using a multichannel pipet.
5. Seal the plate with cap strips.
6. Centrifuge at 1000g for 10 s.
7. Place in a thermal cycler (*see Note 19*) and select program (*see Note 20*); an example program is given in **Fig. 2**.

3.4. PCR Product Check

For projects that are examining compliant samples, it is possible to proceed directly from the barcode PCR to a sequencing reaction. However, it is often critical to screen PCR reactions for successful amplification products when

Table 1
Master-Mix Recipes for PCR Amplification of the Barcode Region (in μL)

Reagent	1 Reaction	Plate
10% Trehalose	6.25	650
ddH ₂ O	2	208
10 \times PCR buffer	1.25	130
50 mM MgCl ₂	0.625	65
10 mM dNTP	0.0625	6.5
10 μM Primer 1	0.125	13
10 μM Primer 2	0.125	13
<i>Taq</i> polymerase	0.0625	6.5
Mix volume	10.5	-
DNA template	2	-
Total volume	12.5	-

working with older specimens or with a new taxonomic group. This has traditionally been a laborious task involving gel casting and the loading of individual reaction products. The use of bufferless, precast agarose gels circumvents these and other problems (*see* **Note 21**).

1. Plug the Mother E-Base into an electrical outlet. Press and release the “pwg/prg” (power/program) button on the base to select program EG. Select a run time by pressing the “time” button (*see* **Note 22**).

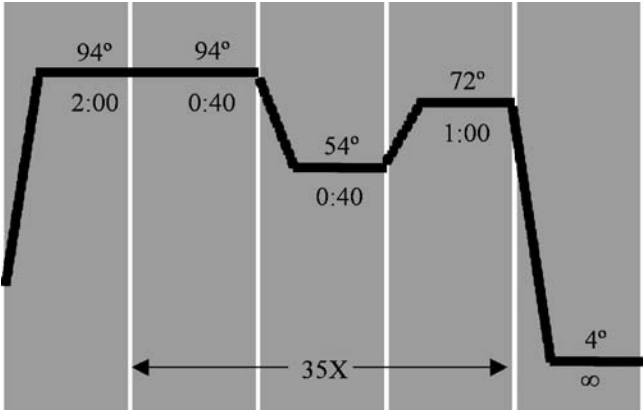


Fig. 2. A sample thermocycler program for PCR amplification of the COI barcode region.

Table 2
Primers for the PCR Amplification of COI for Varied Taxonomic Groups

Primer name	Sequence (5'-3')	Taxonomic group	Reference
LCO1490	GGTCAACAAATCATAAAGATATTGG	Misc. Phyla	32
HCO2198	TAAACTTCAGGGTGACCAAAAAATCA		
BirdF1	TTCTCCAAACCACAAAGACATTGGCAC	Birds	5
BirdR1	ACGTGGGAGATAATTCCAAATCCTGG		
FishF1	TCAACCAACCACAAAGACATTGGCAC	Fishes	7
FishR1	TAGACTTCTGGGTGGCCAAAAGAAATCA		
FishF2	TCGACTAATCATAAAGATATCGGCAC		
FishR2	ACTTCAGGGTGACCCGAAGAATCAGAA		
VF1	TTCTCAACCAACCACAAAGACATTGG	Mammals, reptiles, Amphibians	7, 33
VF1d	TTCTCAACCAACCACCAARGAYATYGG		
VF1i	TTCTCAACCAACCACCAIGAIAITIGG		
VR1	TAGACTTCTGGGTGGCCAAAAGAAATCA		
VR1d	TAGACTTCTGGGTGGCCRAARAAYCA		
VR1i	TAGACTTCTGGGTGICCIAIAAICA		
LepF1	ATTCAACCAATCATAAAGATATTGG	Insects	17
LepR1	TAAACTTCTGGATGTCCAAAAAATCA		
CrustF1	TTTTCTACAAATCATAAAGACATTGG	Crustaceans	9, 32
CrustF2	GGTTCTTCTCCACCAACCACAAARGAYATHGG		
HCO2198	TAAACTTCAGGGTGACCAAAAAATCA		
Chel F1	TACTCTACTAATCATAAAGACATTGG	Chelicerates	10
Chel R1	CCTCCTCCTGAAGGGTCAAAAAAATGA		
Chel R2	GGATGGCCAAAAAATCAAAATAAATG		

2. Remove the gel from the package and remove the plastic comb from the gel. Slide the gel into the two electrode connections on the Mother E-Base.
3. Load 16 μL of ddH₂O with a multichannel pipettor.
4. Load appropriate DNA markers in the marker wells if necessary.
5. Load 4 μL of the sample with a multichannel pipettor.
6. To begin electrophoresis, press the “pwd/prg” button. The red light should change to green.
7. At the end of run (signaled with a flashing red light and rapid beeping), press and release the “pwr/prg” button.
8. Remove the gel cassette from the base and capture a digital image of the gel with a gel documentation system.
9. Arrange lanes and manipulate image as necessary using the Invitrogen E-Editor software (available at www.invitrogen.com/egels).
10. Incorporate the E-gel image into an “electronic lab book” spreadsheet for hit picking (*see* **Note 23**). (See www.barcodeoflife.org for lab spreadsheets and documentation.)
11. Gels can be stored at 4°C for at least one reuse.

3.5. Sequencing Setup

Cycle sequencing reactions are set up directly from PCR products (*see* **Note 24**), minimized to cut costs (*see* **Note 25**), and can be premixed and stored frozen for convenience and quality control measures (*see* **Note 26**).

1. Defrost your reagents and place in a cold block.
2. Mix reagents in a 1.5-mL tube following the recipe given in **Table 3** (*see* **Note 27**). Prepare a mix for both the forward and reverse reaction of each sample.
3. Vortex-mix gently and aliquot 9 μL of the sequencing mix into each well of the microplate. This can be done quickly by aliquoting 1/8 (~117 μL) of the total mix into an eight-tube PCR strip and dispensing to the microplate with a multichannel pipet.

Table 3
Master-Mix Recipes for Sequencing Reaction Setup (in μL)

Reagent	1 Reaction	Plate
Dye terminator mix v3.1	0.25	26
5× Sequencing buffer	1.875	195
10% Trehalose	5	520
10 μM Primer	1	104
ddH ₂ O	0.875	91
Mix volume	9	—
PCR product	1	—
Total volume	10	—

4. Add 0.5–2 μL of PCR product (*see Note 28*) to each well using a multichannel pipet.
5. Seal the plate with cap strips.
6. Centrifuge at 1000g for 10 s.
7. Place in a thermal cycler and select the program (*see Note 29*); a sample program is given in **Fig. 3**.

3.6. Sequencing Reaction Cleanup

The Sephadex column method of cycle sequencing reaction cleanup is a rapid, reliable, and cost-effective alternative to both traditional and newer cleanup methods (*see Note 30*).

1. Use the column loader to fill each well of the filter plate with Sephadex.
2. Hydrate the wells with 300 μL of ultrapure H_2O .
3. Let the Sephadex swell before use, either overnight in the refrigerator or for 3–4 h at room temperature.
4. Drain the excess water by using the centrifuge alignment frame to attach the Sephadex filter plate to an empty microplate. Rubber bands can be used to hold the plates in place. Centrifuge at 750g for 3 min and discard the water.
5. Using a multichannel pipet, add the sequencing reactions ($\sim 10 \mu\text{L}$) to the center of the Sephadex columns in each well.
6. Using a multichannel pipet and an eight-tube PCR strip, add 10 μL of formamide to each well of a sterile 96-well reaction plate.
7. To elute the purified sequencing reactions into the formamide, attach the reaction plate to the bottom of the Sephadex filter plate, securing it with rubber bands. Centrifuge at 750g for 3 min.

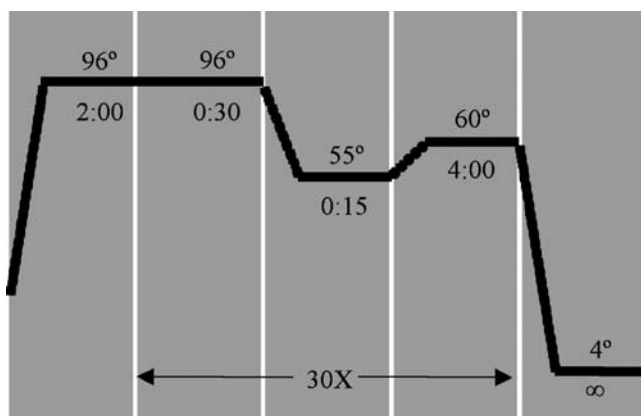


Fig. 3. Thermocycler program for sequencing reactions.

3.7. Sequence Analysis

The high production goals of a DNA barcoding facility make a multiple capillary instrument essential. Applied Biosystems and Amersham produce several highly reliable instruments with varied production capacities (*see Note 31*). The following protocol refers to an Applied Biosystems 3730 capillary sequencer.

1. Cover the reaction plate with septa.
2. Place the reaction plate into the plate base and attach the plate retainer.
3. Print and affix a barcode to the assembled plate.
4. Stack assembled plate in the 3730.
5. Perform routine 3730 maintenance as necessary.
6. Using the Plate Manager of the Data Collection software (Applied Biosystems), import the plate record(s) for the plate being run.
7. Begin the run within Run Scheduler.

3.8. Sequence Editing/Alignment

The minimum capabilities of a sequence editing and alignment software package for high-volume barcoding are the abilities to assemble bidirectional reads and edit trace files. Sequencher, SeqScape, and Lasergene are capable, plus provide other useful functions.

1. Using the sequence alignment editor of choice, open up the trace files of the finished run. Depending on the software being used, as meant samples can be analyzed at once.
2. Assemble the forward and reverse read of each sample.
3. Remove the primer sequences from the end of the read.
4. Review each base call, making adjustments to miscalled or uncalled bases as necessary.
5. Save sequence(s) in Fasta format and upload to the appropriate project in BOLD or to another online sequence repository.

4. Notes

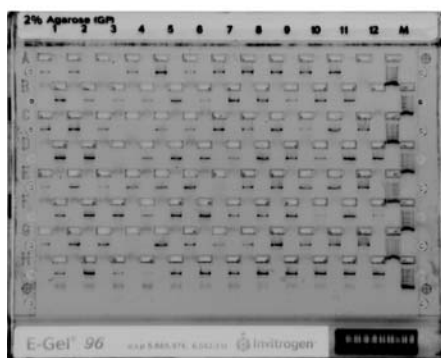
1. Killing in a DNA-friendly fashion refers to freezing, cyanide exposure, or immersion in ethanol, and avoiding even brief exposure to killing/preservation agents such as ethyl acetate or formalin that damage DNA. DNA in dried specimens ordinarily remains in good condition for at least a year, but degradation becomes increasingly problematic as time passes. DNA in frozen specimens (especially those held in cryogenic conditions) remains stable indefinitely, but DNA in ethanol-preserved material often degrades as a result of acidification. As a result, barcode analysis should follow collection as quickly as possible.
2. All samples should be handled on a clean working surface and all tissue-handling instruments should be acid or flame sterilized between each sample. A Bunsen

burner flame is convenient for sterilization; small propane tanks are ideal for settings where gas is not online.

3. In any laboratory that seeks high production rates, it is critical to carry out all stages of barcode analysis in 96-well plates. Two wells should be left blank for positive and negative controls in the preceding steps, leaving 94 wells for samples. Care must be taken to avoid cross-contamination between wells when loading these plates with samples.
4. In addition to the Chelex and membrane-based methods discussed here, there are many alternate kits and protocols for DNA release or isolation. The use of magnetic beads is becoming increasingly popular, primarily because of their amenability to automation. When speed is critical, several kits offer DNA extraction in minutes, such as the Extract-N-Amp PCR Kit (Sigma-Aldrich) and FTA cards (Whatman, Florham Park, NJ). These methods are currently being evaluated for their value in high-throughput DNA barcoding.
5. DryRelease is a Chelex-based, DNA release protocol that rapidly liberates DNA into solution, making it accessible for downstream applications (24). It requires minute tissue samples and minimal technician time, but is not suitable for samples with high levels of PCR inhibitors (e.g., hemoglobin), for samples in which DNA is degraded, or where pure DNA for long-term storage is required.
6. Working volumes can range from 30 to 110 μL depending on sample size. For example, small crustaceans or legs of small insects should be extracted in 30 μL , whereas small blocks of vertebrate tissue can be extracted in 100–110 μL .
7. The membrane-based DNA extraction method of the Nucleospin 96 Tissue kit relies on DNA binding to the silica membrane in the presence of a high concentration of chaotropic salt. It results in highly pure DNA and is exceptionally sensitive (25), making it useful for studies on specimens with degraded DNA (24). The higher cost and large demands on technician time of this and similar methods can limit its utility.
8. A manual or electronic multichannel pipet is required to effectively perform 96-well DNA extractions with this and most other kits. Similarly, nearly all steps of 96-well kits and protocols can be automated on robotic liquid handling stations if available.
9. As with DryRelease, elution for smaller specimens (or specimens where DNA is likely degraded) should be done with a small volume of water, whereas larger volumes should be employed for larger or fresh specimens.
10. Centrifuging the plates on a square-well block or an open rack of MN tube strips prevents the wells of the microplate from being crushed or cracked by the force.
11. Primer design is critical and minor adjustments can have large impacts on barcode recovery. The first phase of any study on a new group should involve a serious effort to identify optimal primers. To address mismatches between primers and target DNA, the use of degenerate or inosine containing primers is recommended (26). Primers with two to four degenerate positions or inosine bases will often rescue barcodes from recalcitrant specimens and may also protect from amplifying nuclear pseudogenes (27).

12. The use of sterile barrier tips is recommended for all PCR reagents to avoid contamination. Clean the bench top with alcohol or detergent before setting up reactions. DNA templates (DNA extracts or PCR products) should be kept away from the PCR reagents while you are setting up the reaction mixes. Add DNA only after all of the reagents have been returned to the freezer.
13. The addition of trehalose is optional but may enhance PCR success (28) (see Fig. 4) and makes it possible to freeze aliquoted master mixes (29). Master mix can be stored in tubes at -20°C for 1–3 months or aliquoted directly into 96-well plates at -20°C for up to 1 month.
14. To reduce costs, reaction volumes can be significantly lowered. To dispense such small volumes accurately, it is useful to make up enough master mix for several plates. Note that it is necessary to include extra volume to allow for pipetting mistakes and dead volume in the multichannel pipet (e.g., to make ten 96-well plates with 12.5- μL reactions per well, include about 40 extra reactions).
15. The presence of PCR inhibitors can usually be overcome by incorporating amplification facilitators such as trehalose, bovine serum albumin (BSA), betaine, or dimethyl sulfoxide in the PCR mix (28,30,31).
16. There is a growing diversity of polymerases or polymerase cocktails. Some enable PCR to be executed much more quickly (e.g., Z Taq, Takara Bio, Otsu, Japan); others aid the amplification of damaged templates (Restorase, Sigma-Aldrich) or permit high-fidelity replication (Diamond Taq, Bioline, Randolph, MA).
17. The amount of DNA extract used will depend on the specimen and extraction method employed. Although this may take some adjustment, it is not usually

A)



B)

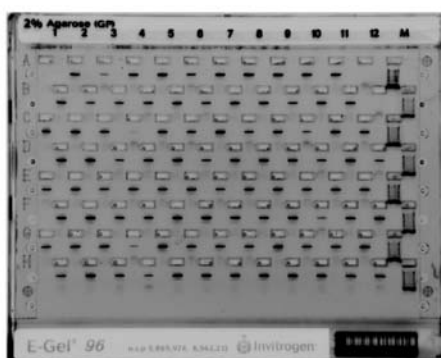


Fig. 4. Demonstration of PCR-enhancing ability of the additive trehalose. A dark band on the E-gel images (negative exposure) indicates a successful PCR amplification for that sample; the clear slots indicate the loading wells. A12 and B12 are negative controls; column *M* contains size markers. (A) Regular PCR master mix without trehalose, and (B) PCR master mix with 5% trehalose.

necessary to quantify genomic DNA extracts because even a few copies of the target gene are sufficient for PCR amplification. It is best to keep the volume of DNA template as low as possible to avoid adding reaction inhibitors that may be present, and to avoid unspecific amplification.

18. Always include a sample without template as a negative control to check for contamination of the reagents. Also include a positive control (a DNA sample that has amplified in the past) to test the effectiveness of the PCR reagents.
19. The latest generation of thermal cyclers (e.g., Eppendorf MasterCycler EP Silver) have faster thermal ramping that allows PCR amplifications to be completed more quickly (1–2 vs. 3–4 h).
20. The annealing temperature for a PCR reaction is generally the only variable that needs to be changed when using new primers. Setting the annealing temperature at 2–6°C lower than the melting temperature of your primers is a general rule of thumb.
21. The Invitrogen E-gel 96 system is quick, sensitive, consistent, and minimizes exposure to ethidium bromide. The capital cost is low, the gels are moderately priced, and the technical time is negligible. Bio-Rad and Amersham Biosciences have similar products.
22. For an approx 700-bp amplicon, it takes only 6 min to fully resolve the band.
23. The use of electronic lab books with incorporated E-gel images facilitates the preparation for hit picking of successful PCR reactions. This is a difficult and time-consuming task; a liquid handling system should be used if available.
24. PCR products are often purified to remove unincorporated nucleotides and residual primers. If this step is omitted, it leads to degradation in the sequencing results for the first 20–50 bp. Such degradation is of little concern when the PCR product is slated for bi-directional sequencing. However, when the PCR product is sequenced in just a single direction, several protocols (e.g., ethanol precipitation) and numerous kits are available such as MultiScreen Filter Technology (Millipore).
25. To reduce costs, sequencing reactions can be cut to 10 µL and prepared with just 0.25 µL of BigDye (1/16 concentration). Because BigDye is one of the costliest reagents in the barcoding protocol, lowering its usage is a critical step in minimizing costs.
26. Sequencing reaction cocktails can be prepared in either 96-well plates or in larger volume tubes well in advance of use, then frozen for up to 3 months. Trehalose is added to ensure stability of the enzyme during freeze–thaw cycles.
27. BigDye v.3.1. cycle sequencing chemistry provides a robust sequencing platform, consistently producing long (~750 bp), high-quality reads, even on GC-rich templates. Amersham Biosciences provides a creditable alternative with the DYEnamic ET Terminator cycle sequencing kit that is fully compatible with Applied Biosystems instruments.
28. The volume of PCR template to add should be estimated based on the intensity of bands on the PCR check gel.

29. The annealing temperature of the program may be varied according to the primer specificity, but 55°C works well for most COI barcode sequencing reactions.
30. Many high-volume genomics facilities use either ethanol precipitation or magnetic bead protocols for sequencing reaction cleanup. Solid-phase reversible immobilization (SPRI) based methods may be particularly well suited to high-throughput barcoding labs.
31. The Applied Biosystems multicapillary sequencer models include the 3100, 3130, 3730, and 3730XL with 4, 16, 48, and 96 capillaries, respectively. Similarly, the Amersham Biosciences line includes the MegaBASE 500, 1000, and 4000 with up to 48, 96, and 384 capillaries, respectively.

Acknowledgments

We thank Alex Borisenko, Rob Dooh, Teresa Crease, Robert Hanner, Angela Holliss, Stephanie Kirk, Paula Mackie, Pia Marquart, Erin Penton, Keith Pickthorn, Cadhla Ramsden, Sujeevan Ratnasingham, Alex Smith, Janet Topan, Taika von Konigslow, Adam Yule, and Tyler Zemlak for their aid in protocol development. Grants from the Gordon and Betty Moore Foundation, NSERC, the Canada Foundation for Innovation and the Ontario Innovation Trust aided preparation of this chapter.

References

1. Hebert, P. D. N., Cywinska, A., Ball, S. L., and deWaard, J. R. (2003) Biological identifications through DNA barcodes. *Proc. R. Soc. B* **270**, 313–322.
2. Blaxter, M. (2003) Counting angels with DNA. *Nature* **421**, 122–124.
3. Hebert, P. D. N., and Gregory, T. R. (2005) The promise of DNA barcoding for taxonomy. *Syst. Biol.* **54**, 852–859.
4. Savolainen V., Cowan, R. S., Vogler, A. P., Roderick, G. K., and Lane, R. (2005) Towards writing the encyclopaedia of life: an introduction to DNA barcoding. *Philos. Trans. R. Soc. B* **360**, 1805–1811.
5. Hebert, P. D. N., Stoeckle, M. Y., Zemlak, T. S., and Francis, C.M. (2004) Identification of birds through DNA barcodes. *PLoS Biol.* **2**, 1657–1663.
6. Kerr, K. A. , Stoeckle, M. Y. , Dove, C, Weigt, L. A. , Francis, C. M., and Hebert, P. D. N. (2007) Comprehensive DNA barcode coverage of North American birds. *Mol. Ecol. Notes.* **7**, 535–543.
7. Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., and Hebert, P. D. N. (2005) DNA barcoding Australia's fish species. *Philos. Trans. R. Soc. B* **360**, 1847–1857.
8. Meyer C. P., and Paulay, G. (2005) DNA barcoding: Error rates based on comprehensive sampling. *PLoS Biol.* **3**, 2229–2238.
9. Costa, F. O., deWaard, J. R., Boutillier, J., Ratnasingham, S., Dooh, R. T., Hajibabaei, M., and Hebert, P. D. N. (2006) Biological identifications through DNA barcodes: the case of the Crustacea. *Can. J. Fish. Aquat. Sci.* **64**, 272–295.

10. Barrett, R. D. H., and Hebert, P. D. N. (2005) Identifying arachnids through DNA sequences. *Can. J. Zool.* **83**, 481–491.
11. Hogg, I. D., and Hebert, P. D. N. (2004) Biological identification of springtails (Collembola: Hexapoda) from the Canadian Arctic, using mitochondrial DNA. *Can. J. Zool.* **82**, 749–754.
12. Smith, M. A., Fisher, B. L., and Hebert, P. D. N. (2005) DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Philos. Trans. R. Soc. B* **360**, 1825–1834.
13. Ball, S. L., Hebert, P. D. N., Burian, S. K., and Webb, J. M. (2005) Biological identifications of mayflies (Ephemeroptera) using DNA barcodes. *JNABS* **24**, 508–524.
14. Cywinska, A., Hunter, F., and Hebert, P. D. N. (2006) Identifying Canadian mosquito species through DNA barcodes. *Med. Vet. Entomol.* **20**, 413–424.
15. Smith, M. A., Woodley, N. E., Janzen, D. H., Hallwachs, W., and Hebert, P. D. N. (2006) DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). *Proc. Natl. Acad. Sci. USA* **103**, 3657–3662.
16. Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H., and Hallwachs, W. (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astrartes fulgerator*. *Proc. Natl. Acad. Sci. USA* **101**, 14812–14817.
17. Hajibabaei, M., Janzen, D. H., Burns, J. M., Hallwachs, W., and Hebert, P. D. N. (2006) DNA barcodes distinguish species of tropical Lepidoptera. *Proc. Natl. Acad. Sci. USA* **103**, 968–971.
18. Hebert, P. D. N., deWaard, J. R., and Landry, J-F. (2007) DNA barcodes deliver: species identifications and revelations for 1/1000 of the Animal Kingdom. In preparation.
19. Saunders, G. W. (2005) Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Philos. Trans. R. Soc. B* **360**, 1879–1888.
20. Seifert, K. A., Sampson, R. A., deWaard, J. R., Houbracken, J. A., Levesque, C. A., Montcalvo, J.-M., Louis-Seize, G. and Hebert, P. D. N. (2007) Prospects for fungus identification using CO1 DNA barcodes, with *Penicillium* as a test case. *Proc. Natl. Acad. Sci. USA* **104**, 3901–3906.
21. Kress, J. W., Wurdack, K. J., Zimmer, E. A., Weigt, L. A., and Janzen, D. H. (2005) Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. USA* **102**, 8369–8374.
22. Chase, M. W., Salamin, N., Wilkinson, M., Dunwell, J. M., Kesanakurthi, R. P., Haidar, N., and Savolainen, V. (2005) Land plants and DNA barcodes: short-term and long-term goals. *Philos. Trans. R. Soc. B* **360**, 1889–1895.
23. Marshall, E. (2005) Will DNA barcodes breathe life into classification? *Science* **307**, 1037.
24. Hajibabaei, M., de Waard, J. R., Ivanova, N. V., Ratnasingham, S., Dooh, R. T., Kirk, S. L., Mackie, P. M., and Hebert, P. D. N. (2005) Critical factors for

- assembling a high volume of DNA barcodes. *Philos. Trans. R. Soc. B* **360**, 1959–1967.
25. Hoss, M., and Paabo, S. (1993) DNA extraction from Pleistocene bones by a silica-based purification method. *Nucleic Acids Res.* **21**, 3913–3914.
 26. Batzer, M. A., Carlton, J. E., and Deininger, P. L. (1991) Enhanced evolutionary PCR using oligonucleotides with inosine at the 3'-terminus. *Nucleic Acids Res.* **19**, 5081.
 27. Sorenson, M. D., Ast, J. C., Dimcheff, D. E., Yuri, T., and Mindell, D. P. (1999) Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. *Mol. Phylogenet. Evol.* **12**, 105–114.
 28. Spiess, A. N., Mueller, N., and Ivell, R. (2004) Trehalose is a potent PCR enhancer: lowering of DNA melting temperature and thermal stabilization of *Taq* polymerase by the disaccharide trehalose. *Clin. Chem.* **50**, 1256–1259.
 29. Franks, F. (1990) Freeze drying: from empiricism to predictability. *Cryoletters* **11**, 93–110.
 30. Al-Soud, W. A., and Rådström, P. (2000) Effects of amplification facilitators on diagnostic PCR in the presence of blood, feces, and meat. *J. Clin. Microbiol.* **38**, 4463–4470.
 31. Frankman S., Kobs G., Simpson D., and Storts D. (1998) Betaine and DMSO: enhancing agents for PCR. *Promega Notes* **65**, 27.
 32. Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **3**, 294–299.
 33. Ivanova, N. V., deWaard, J. R., and Hebert, P. D. N. (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Mol. Ecol. Notes* **6**, 998–1002.