Chapter 15

DNA Mini-barcodes

Mehrdad Hajibabaei and Charly McKenna

Abstract

Conventional DNA barcoding uses an approximately 650 bp DNA barcode of the mitochondrial gene COI for species identification in animal groups. Similar size fragments from chloroplast genes have been proposed as barcode markers for plants. While PCR amplification and sequencing of a 650 bp fragment is consistent in freshly collected and well-preserved specimens, it is difficult to obtain a full-length barcode in older museum specimens and samples which have been preserved in formalin or similar DNA-unfriendly preservatives. A comparable issue may prevent effective DNA-based authentication and testing in processed biological materials, such as food products, pharmaceuticals, and nutraceuticals. In these cases, shorter DNA sequences—mini-barcodes—have been robustly recovered and shown to be effective in identifying majority of specimens to a species level. Furthermore, short DNA regions can be utilized via high-throughput sequencing platforms providing an inexpensive and comprehensive means of large-scale species identification. These properties of mini-barcodes, coupled with the availability of standardized and universal primers make mini-barcodes a feasible option for DNA barcode analysis in museum samples and applied diagnostic and environmental biodiversity analysis.

Key words: DNA barcoding, Museum specimens, Biodiversity, COI, Formalin, Molecular diagnostics

1. Introduction

DNA barcoding is a cost-effective genomics tool for species identification and a valuable approach to aid the discovery of new and cryptic species (1, 2). Since its inception, the Barcode of Life initiative has revolutionized biodiversity assessment and has significantly aided in fields, such as bio-security, cryptic species identification, paleocological studies, forensics and diet analysis (3–6)

The 650 bp animal DNA barcode—near the 5′ region of the mitochondrial cytochrome *c* oxidase 1 (COI, *cox1*) gene (1)—is readily sequenced and provides species-level resolution of approximately 98% for large taxonomic assemblages, such as birds, mammals, fishes, and various arthropods (7). Current barcode library construction has mainly focused on the analysis of recently collected

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_15, © Springer Science+Business Media, LLC 2012

specimens (i.e., up to 5 years old) or samples that have been preserved in a DNA-friendly manner (3). While these samples provide an efficient means to obtain cost-effective barcode sequences, it is crucial to obtain genetic information from millions of older museum specimens. There are a number of advantages associated with the use of museum collections; the avoidance of costly field collections, the availability of rare or extinct taxa (5) and a means to build broad geographical barcode libraries (3). Most importantly, museum samples represent the vast and valuable taxonomic knowledge that has been generated by analyzing these specimens. Hence, obtaining genetic information from these well-studied specimens will enable an efficient way of linking available taxonomic knowledge to recently collected specimens through the use of DNA barcodes. For example, museum specimens include type specimens that are considered to be a gold standard for the taxonomic information linked to a species. Therefore, recently collected and barcode identified individuals from presumably the same species, can only be verified if a DNA barcode sequence from the type specimen is available (4).

Conventional barcoding methodology is often limited by its failure to amplify and sequence degraded DNA, which is often found in museum specimens and in preserved and processed biological material (food products, decayed tissues (4)). Approaches to repair DNA in vitro are inefficient and not cost-effective as the DNA damage and degradation in museum samples is complex and difficult to characterize (6, 8). In comparison, short sequences (i.e., 100 bp) are usually stable in museum specimens (6). The use of a short or minimalist barcode (100–300 bp, referred herein as "mini-barcode") greatly expands the applications of DNA barcoding (3) (Table 1).

Most specimens of archive collections have not been assembled with the purpose of genetic studies. The variance in DNA quality

Table 1
Comparison of full-length DNA barcode and mini-barcode

	Full-length barcode (650 bp)	Mini-barcode (100–300 bp)
Specimen sequence success relative to age	>90%, 5–10 years	>90%, up to 200 years
Species resolution	95–98%	91–95%
Technology	Sanger (ABI)	Sanger (ABI) NextGen sequencing (i.e., 454) Single pyrosequencing (PSQ)
Applicability	Barcode library construction Routine barcoding	Museum and preserved samples Processed material (i.e., food products, pharmaceuticals) Environmental barcoding

among similar aged samples can be largely accounted for by the variation in preservation methods. Many specimens, particularly insects, are pinned for storage allowing the soft tissue to desiccate and decompose (9). This dehydration combined with exposure to the environment can lead to a reduction of DNA quality and increased incidences of DNA fragmentation (6). Field collected specimens and pathology tissue samples are commonly exposed to formaldehyde, which has severe consequences on the DNA (6). In a formalin-exposure study utilizing amphibian specimens, a clear negative correlation was found between exposure time and PCR success (6). Furthermore, a recent study by Baird et al. (10), examined the effect of formalin preservation on DNA barcoding. This work used tissue samples of four invertebrate species commonly used in freshwater biomonitoring programs as well as archival specimens of macroinvertebrates. The authors concluded that exposure to formalin followed by long-term storage can dramatically reduce the ability to obtain a full-length barcode; however, some mini-barcodes can still be recovered from these samples (10). In many cases, damaged DNA can be extracted from a sample but the DNA is broken into small fragments due to hydrolysis of the DNA backbone (6, 11). This fragmented DNA is unable to be amplified using standard barcoding primers. Models derived from preserved specimen and tissue samples by capillary electrophoresis predict a rapid initial decline in average DNA fragment size in the first 5 years followed by a more gradual change in the following time period (6). The same problems that plague preserved samples affect fossils and ancient DNA. Our understanding of evolutionary processes is hindered due to DNA fragmentation, cross-linking due to condensation (12) and pyrimidine oxidation which prevents extension during the amplification process (11). Furthermore, these ancient records often produce DNA extracts that are a combination of bacterial, fungal, and human contaminants (11) complicating the ability to achieve a useable standard barcode.

Mini-barcodes (e.g., 100–300 bp) have been found effective for species-level identification in DNA-damaged samples and in situations, where it is difficult to obtain a full-length barcode (Table 2). Additionally, components, such as average nucleotide composition, patterns of strand asymmetry, and a high frequency of hydrophobic amino acid encoding codons can be accurately predicted from a short barcode sequence (13). Furthermore, it has been shown that mini-barcodes may provide measures at both the intra-specific and intra-generic levels of sequence variability and divergence in some cases when compared to full length barcodes (3). Full-length 650 bp COI barcodes can exhibit up to 98% species resolution, with smaller regions 100 bp and 250 bp producing correspondingly lower rates of identification success (3, 4) (Fig. 1), but when employed in ecological or environmental contexts where the number of species per genus is often low, they can produce

Table 2
Different research projects that used DNA mini-barcodes for biodiversity analysis and species identification

Sample type	Age of sample (years, unless specified)	Taxonomic group	Gene	Mini-barcode size (bp)	References
Oven-dried (museum)	2–21	Insecta: Lepidoptera	COI	134 and 221	(3)
Ethanol-preserved (museum)	1–14	Insecta: Hymenoptera	COI	135	(3)
Formalin-preservation	>1–23	Insecta: Trichoptera	COI	130	(10)
Arsenic/borax-preserved and air-dried (museum)	Average 80	Aves: 17 orders	COI	190 and 310	(8)
Dried and formalin-fixed tissue (forensic samples)	N/A	Reptilia	C01	175–245	(18)
Museum	Average 15	Insecta: Diptera	COI	227–469	(20)
Silica-dried (museum)	N/A	Plantae	P6 loop of trnH intron	13–158	(15)
Permafrost	22 960 ± 120 (OxA-15348) and 15 810 ± 75 (OxA-14930) uncalibrated Radiocarbon years	Plantae	P6 loop of trnH intron	13–158; average 43.2	(15)
Decayed carcass	N/A	Actinopterygii: Salmoniformes	Ots213	170–375	(17)

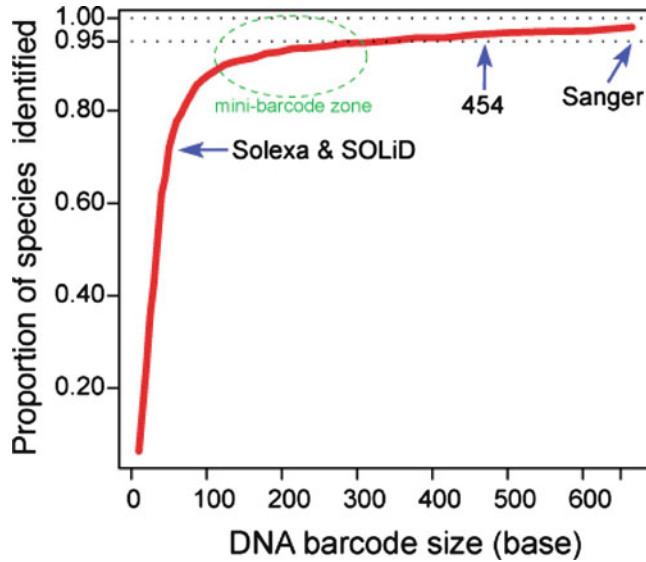

Fig. 1. Comparison of DNA barcode size versus proportion of species identified reveals the efficiency of mini-barcodes in resolving species (adapted from ref. 4). Sequence read lengths typically obtained from three commonly used next-generation sequencing technologies as well as Sanger sequencing are shown on the graph. It is clear that 454 pyrosequencing and Sanger are currently optimal technologies for mini-barcode and full-barcode recovery.

rates of identification that are very high (4). In silico studies have been utilized to corroborate the empirical tests of the rates of identification success for DNA barcodes, but also point to the need to carefully design experiments in environmental contexts where primer bias may affect the results (3, 14).

This discovery has led to an increase in the use of mini-barcodes. The consistency of mini-barcodes to distinguish between species has been explored in plants (15), fish (3, 4, 14, 16, 17), reptiles (18), birds (4, 8, 19), arthropods (3, 4, 9, 20–23) fungi (4, 20), and mammals (4, 14). Additionally, multiple overlapping mini-barcodes have been used to reconstruct the full COI barcode (9, 19). Applications of mini-barcodes include food Web analysis (21), distinction of cryptic species (16), biodiversity studies (3, 11, 15, 17, 22, 23), and effective law enforcement for the conservation of wildlife (18).

Short DNA regions can also be utilized via new parallel high-throughput sequencing platforms (aka next-generation sequencing), such as pyrosequencing-based 454 Roche sequencer allowing a comprehensive and inexpensive means for barcoding applications (15, 23) (Fig. 1). With these technologies, the need for traditional cloning is eliminated as simultaneous amplification of several thousands to millions of 10–400 bp DNA molecules is achieved during the emulsion PCR process (7).

2. Materials

We recommend using molecular biology laboratory material, including gloves, disposable pipette tips, PCR-grade tubes/strips, or 96-well microtiter plates.

2.1. Silica-Based DNA Extraction

1. NucleoSpin 96 Tissue Kit (Macherey-Nagel).
2. Ethanol (99.9%).
3. Matrix ImpactII P1250 pipette (Matrix Technologies).
4. Centrifuge with deep-well plate rotor (25R, Beckman Coulter).
5. Incubator (Fisher Scientific).

2.2. PCR Amplification

1. 10× PCR Buffer for Platinum Taq DNA polymerase (Invitrogen).
2. 10 mM deoxynucleotide (dNTP) mix (New England Biolabs).
3. Oligonucleotide primers (Forward and Reverse), 100 mM stock, and 10 mM working solutions (Integrated DNA Technologies).
4. Platinum Taq DNA Polymerase (Invitrogen).
5. 50 mM Magnesium chloride ($MgCl_2$) (Sigma-Aldrich).
6. Molecular biology grade distilled water.
7. Thermocycler (Mastercycler EP Gradient, Eppendorf).

2.3. PCR Amplification Check Using E-gel 96 Gel

1. Mother E-base (Invitrogen).
2. 2% E-gel 96 gels (Invitrogen).
3. Gel documentation system (AlphaImager 3400, Alpha Innotech Corporation).

2.4. PCR Amplification Check Using Handmade Agarose Gel

1. Molecular biology grade Agarose (Sigma-Aldrich).
2. 1× TBE buffer: 0.9 M Tris base, 0.89 M Boric acid, 0.02 M Na-EDTA.
3. Ethidium Bromide (10 mg/ml).
4. DNA size standard ("ladder"; New England Biolabs).
5. Submarine electrophoresis apparatus and power supply (Thermo EC, Fisher Scientific).
6. Gel documentation system (AlphaImager 3400, Alpha Innotech Corporation).

2.5. Sanger Sequencing Reaction

1. BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems).
2. 5× Sequencing Buffer: 400 nm Tris–HCl pH 9.0 and 10 mM $MgCl_2$.

Table 3
PCR Primer sets commonly used for mini-barcode amplification

Name	Primer sequence (5′–3′)	Target	Amplicon size (bp)	References
Uni-MinibarF1	TCCACTAATCACAARGATATT GGTAC	COI; Universal eukaryotes	130	(4)
Uni-MinibarR1	GAAAATCATAATGAAGGCATGAGC			
Minibar-F1	TGA TTY TTT GGH CAC CCR GAA GT	COI; Reptile, snake	175	(18)
Minibar-R1	AAT ATR TGR TGG GCY CAD AC			
Minibar-F2	GGT AGY GAT CAA ATC TTT AAY GT	COI; Reptile, snake	245	(18)
Minibar-R2	GGG TAG ACD GTT CAV CCT GTT CC			
Primerpair-1	AGCATTAGCTCTCCCTGA AGCCATACGGCGGTGAAT	16S; plant	166	(15)

3. Sequencing Oligonucleotide Primers. For bi-directional sequencing, two sequencing reactions are performed, each require a single Forward or Reverse primer. See Table 3 for a list of primers.

3. In Silico Approaches for Mini-barcode Analysis

3.1. Mini-barcode Primer Selection

Primers should be designed based on an alignment of sequences from a reference barcode library source, such as GenBank or BOLD (http://www.barcodinglife.com/views/idrequest.php). Keeping in mind physical and structural properties, such as G+C content, annealing temperature and self complementary, oligos should target highly conserved regions flanking stretches of 100–300 bp in barcode region. To facilitate high-throughput sequencing applications, M13 tails may be attached to the forward and reverse primers. Although the addition of these tails generally do not bias PCR results, it is best to verify this empirically before using tailed primers in real applications (4). Primer suitability can be confirmed using a number of freely available programs, such as Primer3 (24) and IDT OligoAnalyzer (25). Most applications of mini-barcodes allow for a selection of a target taxonomic group and subsequent primer design. For example, Hajibabaei et al. (3) developed primers for mini-barcoding specific Lepidoptera species. In this case, the availability of reference sequences from closely related taxa to the target species will increase the chances of developing robust

primers. For protein-coding genes, such as COI, conservation of primer binding sites at amino acid level will aid the selection of target primers. However, when dealing with unknown specimens, or in cases where identification speed is required, universal primers for mini-barcodes can be utilized. Such a universal mini-barcode system has been developed for COI and tested on a number of taxonomic groups (4). Additionally, universal primers are important for sequencing mini-barcodes from environmental samples containing different organisms (23). Table 3 summarizes different commonly used primers for mini-barcoding.

3.2. Bioinformatics Analysis to Estimate Mini-barcode Performance

Before using a mini-barcode in empirical tests, it is important to determine whether sequence information obtained from a given mini-barcode sequence can provide species-level resolution in a given taxonomic group. Available taxon-specific barcode data can be downloaded from barcode libraries, such as BOLD (http://www.barcodinglife.com/views/idrequest.php) or GenBank. Alternatively, sequences obtained in full-length barcode analysis in a given taxonomic group can be used as template for primer design for mini-barcoding. The full-length region can then be divided into short subsets from 5′ to 3′ end (for example see Subheading 3). Software such as MEGA (26) provides a simple bioinformatics tool to partition data for this purpose. By comparing each putative mini-barcode segment (i.e., the first 100 bp of 5′ region) to the full-length barcode, through simple statistics, such as number of variable and parsimony informative sites and intra-specific and intra-generic divergences (3) one can determine the mini-barcode fragment with optimal information. Subsequent neighbor-joining (NJ) analysis can measure resolution and ultimately tests the accuracy of the short DNA fragments for the practice of species identification for the target taxonomic group (3, 7).

3.3. Testing the Performance of a Putative COI Mini-barcode for Species Identification

1. Select a target sequence library from BOLD or GenBank or use a set of taxonomically identified sequences. Typically, 100–500 taxa are optimal for this analysis (see Note 1).
2. Align sequences using automated alignment tools (ClustalW, MUSCLE). These tools are available stand-alone or embedded in sequence analysis software, such as MEGA.
3. Inspect alignment visually using a program, such as MEGA. Look for obvious signs of misaligned sequences, such as indels (insertions/deletions). Use amino acid translation to guide the alignment verification.
4. Save the aligned sequences in Mega format (.meg). If you use MEGA for alignment, this can be done from alignment viewer. If you use other programs for alignment, save your alignment in FASTA format (.fasta, .fas) and import the file in MEGA.

5. Open the file for phylogenetic analysis in MEGA. Use "Sequence Data Explorer" tool in MEGA to select mini-barcode fragments for analysis. This can be done in "Data" using "Select & Edit Genes/Domains".
6. Select your putative mini-barcode(s) from the full-length library by specifying the beginning and the ending nucleotides. Make sure these positions do not include positions of your forward and reverse primers. You can select multiple mini-barcodes as long as the positions are not overlapping. For overlapping mini-barcodes, you will require the use of a separate analysis for each overlapping mini-barcode.
7. Select your target mini-barcode by checking the box next to it in the same menu (Select & Edit Genes/Domains). Once done, check the alignment in "Sequence Data Explorer" to ensure only your target mini-barcode is highlighted.
8. Use "Highlight" menu in "Sequence Data Explorer" to highlight/measure Variable and Parsimony Informative sites as simple statistics. These numbers can be compared with numbers obtained from a full-length barcode or other mini-barcodes.
9. Use "Phylogeny" tools to assemble an NJ tree for your target mini-barcodes. Use this tree to inspect species-level resolution. Compare this NJ tree to the one obtained from full-length barcode library. Evaluate cases where the mini-barcode does not provide species-level resolution. If these cases are among your putative target taxa for mini-barcoding, it is important to consider a longer or alternative fragment for mini-barcoding. Note that DNA degradation in museum samples often dramatically decreases the chances of obtaining fragments longer than 150–250 bp.
10. Once your optimal mini-barcode fragment(s) is selected, proceed to design forward and reverse PCR primers using 20–30 bases flanking mini-barcode fragment(s).

3.4. DNA Extraction

Many protocols are available for DNA extraction from a number of different tissue types (see Chapter III. B. on DNA extraction (III) for details). Subsequently, several of these procedures have been incorporated into relatively inexpensive and effective commercial kits, such as the silica-based Nucleospin Tissue Kit (MachereyNagel, Düren, Germany) (see Note 2). Most of our mini-barcode tests have utilized this kit to obtain DNA from specimens (including museum samples). Additionally, a recent study by Shokralla et al. (27) has shown that modern PCR enzymes are capable of amplifying genetic information from preservative ethanol for noninvasive sampling or when no tissue specimen is available. This suggests that DNA extraction may be unnecessary in many protocols and in many cases DNA can be obtained from valuable museum specimens

noninvasively. Because mini-barcodes are much smaller than full-length barcodes, simple DNA extraction and releasing approaches, that may not provide full-length intact barcode DNA, are suitable for mini-barcoding.

3.5. PCR Amplification from Dried Museum Samples

Mini-barcodes can be amplified in a simple PCR reaction containing dNTPs, primers (Forward and Reverse), $MgCl_2$, 10× PCR Buffer, Taq polymerase, and template DNA. A master mix (excluding template DNA) should be prepared for all samples that will be processed. The volume of ingredients is to be calculated based upon the number of samples to be amplified with additional 5% to account for pipette error. No special condition or additive is required for amplification of mini-barcodes and different versions of Taq polymerase should be able to amplify mini-barcode sequences (see Note 3). The thermocycler program will vary depending on the annealing temperature and extension time. For the universal primer set, (4) a touch up PCR program was used: 95°C for 2 min, followed by five cycles of 95°C for 1 min, 46°C for 1 min, and 72°C for 30 s, followed by 35 cycles of 95°C for 1 min, 53°C for 1 min, and 72°C for 30 s, and finally a final extension at 72°C for 5 min (see Note 5). Negative control reactions with no DNA template as well as a positive control reaction should always be included (see Note 4).

To determine PCR success, it is necessary to visualize PCR products on an agarose gel. For high-throughput analysis a precast 2% E-gel 96 agarose (Invitrogen, Burlington, ON, Canada) can be used. Alternatively, casting one's own gel and utilizing a molecular size marker (e.g., 100 bp DNA Ladder, New England Biolabs) can give more comprehensive results. Positive amplification (clean bands) can then be bidirectionally sequenced using standard BigDye chemistry (Applied Biosystems, Foster City, CA) on sequencers, such as a 3730xl DNA analyzer (Applied Biosystems, Foster City, CA, USA). The sequence reads are to be trimmed and edited resulting in clean contigs for phylogenetic analysis using tools, such as CodonCode (CodonCode Corporation) BioEdit (27) and MEGA (25).

1. Select DNA samples for PCR amplification (see Note 2) Calculate the volume of each necessary reagent (see Table 4 for a detailed PCR recipe).
2. Prepare master mix with calculated volumes. It is often easy to use the calculation as a check list to ensure all reagents have been added. Briefly vortex and centrifuge.
3. Briefly spin down Taq polymerase before adding to the master mix. Do not vortex, mix by gentle pipette up and down. Return Taq polymerase to freezer.
4. Dispense master mix to PCR reaction tubes (0.2 ml single tubes, strips, or 96-well plates).

Table 4
A typical PCR recipe for amplification of mini-barcodes

Reagent	Initial concentration	Final concentration	Volume per reaction (μl)
H_2O	–	–	17.5
PCR buffer (Platinum)	10×	1×	2.5
$MgCl_2$	50 mM	2.0 mM	1
dNTPs	10 mM	0.2 mM	0.5
Primer (Forward)	10 μM	0.2 μM	0.5
Primer (Reverse)	10 μM	0.2 μM	0.5
Platinum Taq polymerase	5 U/μl	2.5 U	0.5
Template DNA			2
Final volume			25

5. Add 1–2 μl DNA template to each reaction tube. Ensure that pipette tip is changed each time.
6. Cover PCR tubes securely (caps in the case of strips or seal plates with foil) and label using permanent marker.
7. Centrifuge the plate (or PCR tubes/strip in plate holder) about 1 min at 1,000 × *g* ensuring that the centrifuge is well-balanced.
8. Place tubes, strips, or plate firmly into thermocycler. Double check the covers to minimize evaporation before beginning the required program. Ensure that the thermocycler lid is secure and that the program begins.
9. Perform PCR amplification check using gel-electrophoresis (Protocols 3 and 4) to determine positive samples.
10. (Optional) Use a PCR purification method to clean samples prior to sequencing reaction (i.e., QIAQuick (Qiagen, Duesseldorf, Germany)).

3.6. PCR Amplification Check Using Pre-cast 2% E-gel 96 Agarose (Invitrogen, Burlington, ON, Canada)

1. Remove gel from package and using thumb and reasonable force to remove comb. Use packaging as base for gel while applying sample (see Note 6).
2. Load 14 μl of ddH_2O to wells holding 12-multichannel pipette on slight angle.
3. Load 4 μl of PCR product to wells using 12-multichannel pipette.
4. Slide gel into the electrode connections on E-Base™. Ensure that the E-gel display screen says “EG” and change the time to

appropriate amount (~4 min). Press and release pwr/prg button, the red light should turn to green.

5. Remove gel from base and acquire image using a UV transilluminator and digital camera if necessary.
6. Discard gloves, tips and gel in hazardous waste.

3.7. PCR Amplification Check Using 1.5% Hand-Made Agarose Gel

1. Prepare gel-casting tray with appropriate combs relative to number of samples and secure open edges using gasket system or tape.
2. Weigh 1.5 g agarose powder and add this to 100 ml 1× TBE buffer in a glass beaker, preferably with lid loosely tightened (this calculation may vary for different gel sizes; check your electrophoresis apparatus to verify).
3. Boil mixture in microwave (~2 min) until the powder has completely dissolved and solution is uniform.
4. Allow to cool slightly before adding 3 μl ethidium bromide. Swirl beaker gently to mix (see Note 6).
5. Allow more cooling, you should be able to touch bottom of beaker without burning hand.
6. Steadily pour agarose solution into gel tray, do not move while doing this as it may create bubbles. Ensure that there are no bubbles around combs, if so gently remove these using clean pipette tip.
7. Place gel in electrophoresis chamber filled with 1× TBE buffer, add buffer until gel is fully submerged. Gently remove combs.
8. Load 3–5 μl ladder in the first lane of the agarose gel. Mix by pipette 4 μl PCR product with 2–3 μl loading dye on parafilm or in another plate.
9. Apply PCR samples to gel wells.
10. Connect voltage (100 V) and allow samples to run for 20–30 min. Using UV transilluminator of Geldoc system visualize gel and acquire image.

3.8. Sanger Sequencing Reaction

1. Prepare BigDye terminator (Applied Biosystems, Foster City, CA) master mix to appropriate dilution for size of product (i.e., 1/16 dilution) and according to sample number (see Table 5 for details).
2. Aliquot BigDye mix into PCR tubes, strips, or plate. Add 1.5–2 μl of PCR product as template. If using a plate, securely seal plate using foil or strip caps to prevent evaporation. Perform cycle-sequencing reaction for each primer direction (forward and reverse), an optimized thermocycler protocol can be found on the Canadian Center for DNA Barcoding (CCDB) Web site (http://www.ccdb.ca/) *Protocols: Sequencing*.

Table 5
A typical Sanger sequencing recipe with BigDye (1/16 dilution)

Reagent	One reaction (μl)
Dye terminator mix 3.1	0.25
5× ABI sequencing buffer	1.875
10% trehalose	5
10 μM primer	1
H_2O	0.875
Total volume	9
PCR product	1–2
Total volume	~10

3. Perform cycle sequencing clean-up using method, such as AutoDTR™ 96 (EdgeBio, MD, USA). A detailed protocol can be found on the Canadian Center for DNA Barcoding (CCDB) Web site (http://www.ccdb.ca/) in *Protocols: Sequencing*.
4. After clean-up, submit reactions for sequence analysis using an automated DNA sequencer (e.g., Applied Biosystems 3730xl DNA Analyzer).

4. Notes

1. Sequence selection is critical as it influences the analysis of the utility of a putative mini-barcode. For barcoding purposes, species-level discrimination is most important. Hence, sequences used for mini-barcode selection should include maximum number of species. Congeneric species are good targets for this analysis. Additionally, when possible, multiple sequences from each species should be included so that conspecific variation has been taken into consideration in calculations.
2. Always perform sampling and DNA extraction procedures in a dedicated pre-pcr area. Clean work surface with ethanol or a product, such as ELIMINase Decontaminant. All tissue-handling instruments should be sterilized (preferably by flaming) between samples.
3. A high-fidelity Polymerase, such as Platinum Taq (Invitrogen, Burlington, ON, Canada) is recommended as it requires less optimization and works better with small quantities of template DNA.

4. Always include at least one PCR reaction without template as a negative control to check for reagent DNA contamination. Consider using a positive control (a previously amplified DNA sample) to test the efficiency of the PCR reagents.
5. PCR protocols listed are for a thermocycler with a rapid thermal ramping (e.g., Eppendorf MasterCycler EP). This allows for more efficient annealing and quicker completion of PCR amplification, optimizations will need to be made if a model with slower ramping is used.
6. Ethidium bromide is toxic. Always wear nitrile gloves when utilizing Ethidium Bromide. Discard gloves, tips, and used gels in appropriate hazardous container after usage. Consult MSDS and a laboratory health and safety manual for safe handling/disposal before using.

Acknowledgements

This work was supported by grants from Genome Canada through the Ontario Genomics Institute, Environment Canada and NSERC to MH.

References

1. Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proc R Soc Lond B Biol Sci 270:313–321
2. Hajibabaei M, Singer GAC, Hebert PDN, Hickey DA (2007) DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends Genet 23:167–172
3. Hajibabaei M, Smith MA, Janzen DH et al (2006) A minimalist barcode can identify can identify a specimen whose DNA is degraded. Mol Ecol Notes 6:959–964
4. Meusnier I, Singer GAC, Landry JF et al (2008) A universal DNA mini-barcode for biodiversity analysis. BMC Genomics 9:214
5. Wandeler P, Hoeck PEA, Keller LF (2007) Back to the future: museum specimens in population genetics. Trends Ecol Evol 22:634–642
6. Zimmermann J, Hajibabaei M, Blackburn DC et al (2008) DNA damage in preserved specimens and tissue samples: a molecular assessment. Front Zool 5:18
7. Hajibabaei M, Singer GAC, Clare EL, Hebert PDN (2007) Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. BMC Biol 5:24
8. Patel S, Waugh J, Millar CD, Lambert DM (2009) Conserved primers for DNA barcoding historical and modern samples from New Zealand and Antarctic birds. Mol Ecol Resour 10:431–438
9. Dean MD, Ballard JW (2001) Factors affecting mitochondrial DNA quality from museum preserved Drosophila simulans. Entomol Exp Appl 98:279–283
10. Baird DJ, Pascoe TJ, Zhou X, Hajibabaei M (2011) Building freshwater macroinvertebrate DNA barcode libraries from reference collection material: formalin preservation versus specimen age. J North Am Benthol Soc 30:125–130
11. Poinar HN, Schwarz C, Qi J, Shapiro B et al (2006) Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. Science 311:392–394
12. Evans T (2007) DNA damage. NEB Expressions 2(1):1–3
13. Min XJ, Hickey DA (2007) DNA barcodes provide a quick preview of mitochondrial genome composition. PLoS One 2:e325
14. Ficetola GF, Coissac E, Zundel S et al (2010) An in silico approach for the evaluation of DNA barcodes. BMC Genomics 11:434

15. Sonstebo JH, Gielly K, Brysting AK et al (2010) Using next-generation sequencing for molecular reconstruction of past arctic vegetation and climate. Mol Ecol Resour 10:1009–1018
16. Saitoh K, Uehara S, Tega T (2008) Genetic identification of fish eggs collected in Sendai Bay and off Johban, Japan. Icthyol Res 56:200–203
17. Baumstegier J, Kerby JL (2009) Effectiveness of salmon carcass tissue for use in DNA extraction and amplificaton in conservation genetic studies. N Am J Fish Manag 29:40–49
18. Dubey B, Meganathan PR, Haque I (2010) DNA mini-barcoding: an approach for forensic identification of some endangered snake species. Forensic Sci Int Genet 5:181–184
19. Lee PLM, Prysjones RP (2008) Extracting DNA from museum bird eggs, and whole genome amplification of archive DNA. Mol Ecol Resour 8:551–560
20. Houdt JKJ, Breman FC, Virgilio M, Meyer MD (2009) Recovering full DNA barcodes from natural history collections of Tephritid fruitflies (Tephritidae, Diptera) using mini-barcodes. Mol Ecol Resour 10:459–465
21. Rougerie R, Smith AM, Fernandez-Triana J et al (2010) Molecular analysis of parasitoid linkages (MAPL): gut contents of adult parasitoid wasps reveal larval hosts. Mol Ecol 20: 179–186
22. Smith MA, Fisher BL (2009) Invasions, DNA barcodes and rapid biodiversity assessment using ants of Mauritius. Front Zool 6:31
23. Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ (2011) Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. PLoS One 6:e17497. doi:10.1371/journal.pone.0017497
24. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365–386
25. Owczarzy R, Tataurov AV, Wu Y et al (2008) IDT SciTools: a suite for analysis and design of nucleic acid oligomers. Nucleic Acids Res 36 (web server issue)
26. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599
27. Shokralla S, Singer GAC, Hajibabaei M (2010) Direct PCR amplification and sequencing of specimens' DNA from preservative ethanol. Biotechniques 48:233–234
28. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95–98