Final Report

A Rapid, Efficient, and Effective Assay to Determine Species Origin in Biological Materials

NIJ Grant
#2007-DN-BX-K144

Jared M. Latiolais, Principal Investigator
The Bode Technology Group, Inc.

Abstract

This report details the NIJ-funded research performed by The Bode Technology Group, Inc. ("Bode") over a two year period from July 2008 to August 2010. The primary objective of the research was to develop a single nucleotide polymorphism (SNP) panel for the discrimination of species identity from forensic samples of unknown biological origin; using Mitochondrial DNA (mtDNA) based markers. Bode's approach was to develop proof of concept for the assay by focusing on a small number of ubiquitous species potentially present at North American crime scenes. Using sequence data pulled from NCBI GenBank as well as over 100,000 base pairs of sequence generated on-site, we were able to create gene alignments across all included taxa and species for three mtDNA genes including cytochrome oxidase I (COI), NADH dehydrogenase subunit 4 (ND4), and Cytochrome b (Cytb). The development of canonical SNPs confirmatory for species identity was performed using a neighbor joining phylogenetic algorithm with 10,000 subsequent bootstrap replicates. Potential SNPs were identified by consistency index (CI) values of 1.0 and screened by eye for exclusivity. The resultant SNPs candidates were categorized by gene of origin, species indication, and placement within individual gene sequence. Our intent is for this assay to be used by forensic scientists encountering an unknown biological specimen as an initial screening tool. Using the ABI SNaPshot™ Multiplex kits as the base chemistry for our assay requires multiplexing of both initial polymerase chain reaction (PCR) amplification primers as well as single base extension (SBE) primers. An SBE primer was developed for each of the SNPs for both the forward and reverse strand conformations. Polymerase chain reaction (PCR) primers were then created using the consensus by plurality sequence generated from the alignments to identify conserved areas across all samples. The reactions for each PCR primer were optimized to work across the widest range of species possible then multiplexed into one amplification reaction. This initial reaction provides an ample volume of amplified sample to be used as starting material for SBE reactions, thus it is possible to perform several SBE mini-multiplex reactions from one PCR reaction. The mini multiplexes are plagued by inconsistent representation of species identifying peaks at low template DNA concentrations as well as heightened background noise and the presence of random spurious peaks. A number of methods were attempted to decrease background signal, however they seem to be generated through the use of liberal PCR conditions during the PCR reaction and thus unavoidable. Therefore although SBE primers were spaced so as to enable multiplexing into one single assay; we have tested only the use of three separate mini-multiplexes. To date we have demonstrated the proof of concept for the use of degenerate PCR and multiplexed SBE for the identification of unknown animal samples utilizing three mini multiplexes. However, the continuing presence of random background noise and baseline anomalies deters us from suggesting this method for use as anything other than a research tool. The initial results indicate a presumptive assay with a good deal of promise which needs further development before being ready for widespread use.

Table of Contents

Executive Summary

This is the final report as pursuant to the requirements stated in the deliverables of the two year grant funded by the National Institute of Justice (award number #2007-DN-BX-K144). The work was delayed from the initial start date of January 1, 2008 until June of 2008. Also a no-cost extension of the grant was awarded until 3/31/09 to enable the fulfillment of research objectives. During the course of this project a great deal of progress was made into the identification of canonical SNPs capable of being exploited to identify the presence of unknown biological species to forensic sample mixtures. To this end a survey of publicly available databases, including GenBank, revealed gaps in coverage for commonly encountered animals, such as the common cat and red tail deer. In conjunction with the U.S. Fish and Wildlife Service Forensic Laboratory, additional species were collected, sequenced, and compiled into a database. The full database consists of three mtDNA genes including cytochrome oxidase I (COI), NADH dehydrogenase subunit 4 (ND4), and Cytochrome b (Cytb) across 90 animal species in 10 taxa. By performing a phylogenetic reconstruction using a neighbor joining algorithm, Bode was able to identify taxa specific SNPs which can be exploited in assay form. Each of the SNPs were tested using the ABI SNaPshot^{TM} Multiplex kit, which was chosen based on ease of use and adaptability to thermal cycler and capillary electrophoresis instrumentation already present in many forensic laboratories. Degenerate, "universal", amplification primers were developed, tested individually, screened for efficacy, and multiplexed. Cross-reactivity testing between species was performed to identify problem sets. While individually each of the taxa specific SNPs work appropriately and robustly, the application of liberal thermal cycling parameters for "universal" primers appears to contribute to an elevated downstream signal to noise ratio which

confounds analyses. This work has however proven the concept that canonical SNP primers in multiplex can identify individual contributors to complex mixtures of between two and nine taxa.

Introduction

Over the course of this contract, Bode developed the basis for an assay combining multiplex PCR, multiplex single base extension, and fluorescent detection through capillary electrophoresis that can be used to determine the species of origin in biological materials. The assay simultaneously tests for the presence of Mitochondrial DNA from human and non-human species, as well as the identification of all species contributors to a mixed sample. The PCR-SBE-SNP assay generates a molecular signature composed of a characteristic set of SNPs defining each species or taxa.

The list of taxa identified by this assay includes: Human, *Canis domesticus* (dogs), *Felis catus* (cats), *Mus*(mice), *Bos Taurus* (cows), *Gallus* (chickens), *Cervidae* (deer/elk), *Ursidae* (bears), and *Sus* (pigs). These species were chosen because of their common association with American human households and activities, and for the frequency of their proximate involvement in criminal activities. While some species such as deer, elk, and bear may not seem immediately useful in a panel of this type, at Bode we have come across bones and other tissue from these animals co-mingled with human remains when dealing with mass disasters such as plane crashes. It is also our sincere hope that in the future, a rapid species identification test, such as this, will provide international trade organizations such as CITES with a powerful weapon in the fight against endangered species poaching and trading.

Literature Review

The first forensic use of non-human DNA was in 1993 (Yoon, 1993). Genetic variability of a Palo Verde tree seed pod adhering to a suspect's truck was matched to pods from a Palo

Verde tree at the crime scene (Yoon, 1993). Since 1996, analysis of canine and feline trace evidence has been performed in over 20 criminal investigations (Halverson and Batsen, 2005; Menotti-Raymond et al., 1997; Savolainen and Lundberg, 1999; Shutler et al., 1999). In most of these, short tandem repeat (STR) matches were used to link a victim and suspect through individual identification of a suspect's dog or cat (Halverson and Batsen, 2005; Menotti-Raymond et al., 1997; Savolainen and Lundberg, 1999; Shutler et al., 1999). However, STR-based methods are less useful for species identification and less successful with use of DNA that is degraded or in low concentration. In analyzing compromised samples, mitochondrial DNA (mtDNA) offers a more numerous and robust target. Sequence variation within the mtDNA D-Loop in dogs and wolves has been used as supporting evidence in several criminal investigations (Savolainen and Lundeberg, 1999; Savolainen et al., 2000). Nonetheless, there remains a need for a rapid, sensitive, broad-spectrum single assay to determine species contributors to biological evidence.

Additional markers and methods continue to be tested and evaluated for their utility in forensic contexts. These include STR analysis and direct sequencing of the mitochondrial genome (Bataille et al., 1999; Branicki et al., 2003; Fridez et al., 1999; Herbert et al., 2004; Pfeiffer et al., 2004; Savolainen and Lundeberg, 1999; Savolainen et al., 2000). Variation within the mtDNA D-loop, and the cytochrome b (*Cytb*) and cytochrome c oxidase I (COXI) genes has been tested for efficacy in both human and species identification (Bataille et al., 1999; Branicki et al., 2003; Fridez et al., 1999; Herbert et al., 2004; Pfeiffer et al., 2004; Savolainen and Lundeberg, 1999; Savolainen et al., 2000). STR variation observed within the mtDNA D-Loop of all carnivores was tested, but displayed a significant frequency of heteroplasmy and Fridez et

al. (1999) found little utility for these markers in species identification. Savolainen et al. (2000) obtained similar results testing the repeat region in mtDNA extracted from dog and wolf hair.

Alleles within mtDNA genes are more effective tools for segregating both major clades and sister species. Bellis et al. (2003) used PCR amplification with fluorescently labeled primers to test species differentiation by fragment size from five genes including beta-actin, TP53 tumor suppressor gene, 28s rRNA, the mtDNA D-Loop and *Cytb* markers. Results indicated that little interspecific difference is present for the beta-actin and 28S markers, that the mtDNA markers *Cytb* and the D-Loop provided limited species-specific information, and the TP53 markers displayed the highest discriminatory power.

Pfeiffer et al. (2004) demonstrated that PCR amplification of a segment of the *Cytb* gene followed by restriction digest is an effective tool for species identification, but was limited to discrimination between several non-human animal species and not with closely related taxa (Pfeiffer et al., 2004). Similarly, Bataille et al. (1999) tested a multi-step method of species identification utilizing a duplex PCR amplification of the mtDNA D-Loop and *Cytb*, followed by agarose gel electrophoresis, and sequence analysis. The presence of two bands indicated human origin versus one indicating a non-human contributor. For non-human contributors, sequence analysis of the *Cytb* fragment provided additional discrimination power (Bataille et al., 1999). Branicki et al. (2003) also employed *Cytb* sequence variation as a method of species identification in forensic casework. A total of 34 species were positively analyzed using PCR amplification, cycle sequencing, and GenBank BLAST comparisons. Nonetheless, it was not possible to differentiate between some of the more closely related species solely by sequence variation in *Cytb* (Branicki et al., 2003).

The cytochrome c oxidase I (COXI) gene appears to be more rapidly evolving and amenable to assay development for differentiating closely related species. Herbert et al. (2004) tested the mtDNA cytochrome c oxidase I (COXI) gene and found that interspecific sequence variation in the gene was great enough to segregate all individuals representing 260 species of North American birds. Additionally, intraspecific variation was minimal. Although the mtDNA gene ND4 has not been extensively tested for utility in forensic casework, sequence variation in this gene has been demonstrated to be phylogenetically informative and sufficient for resolving relationships among closely related species of nematodes and turtles (Liu et al., 1999; Blouin et al., 1998).

It should be noted that at the present time, research is being led with regard to identifying an individual animal involved in a crime. The FidoTyping™ assay (Mitotyping Technologies) and MeowPlex assay (Butler et. al., 2002) provide ID for a specific canine or feline involved in a crime using mtDNA and STRs. These tests function like the forensic HID methods currently used for humans. However they do not address the problem of screening biological material with an unknown origin which could be any of a number of different species.

As this review has shown, with some exception, the majority of work performed up to date on forensic species identity is based on sequencing data. The methods previously validated for species identification rely on the use of small gene "barcodes" to function as species tags indicative of entire taxa. This method has been employed to great success; however one can imagine many forensically relevant instances where there would be either human and animal or multiple animal mixtures present. Bode proposes a method whereby species identification can be performed through the exploitation of species specific SNPs that will be rapid, amenable to

current instrumentation present in forensic laboratories, and be able to separate mixtures of species.

Project Strategy

Bode's approach to the building of an animal taxon identification assay is to employ a PCR-SBE-SNP based methodology. After the amplification of a small amount of biological material with multiplexed PCR primers designed to "universally" amplify across all species, a single base extension assay will be employed to identify canonical SNPs within a sample and subsequently identify all taxa present. The single base extension (SBE) technique employs oligonucleotides annealing 5' to 3' to complementary target sequences immediately adjacent to a SNP. Through enzyme-mediated fluorescent-ddNTP addition, the oligonucleotide primer is extended one base. Different alleles of the SNP support incorporation of specific corresponding fluorescent labels and the incorporation of dideoxy nucleotide triphosphate (ddNTP) forces termination of the sequencing reaction following insertion of a single base. In the figure, below, incorporation of ddCTP inserts a fluorescein label (blue-green) while incorporation of ddGTP inserts a rhodamine label (red). These fluorescent tags, representing an original G or C allele of the SNP, respectively, are detected following capillary electrophoresis using a 3100 or 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA)

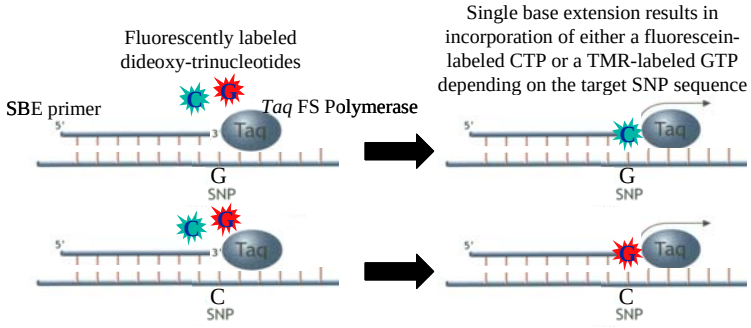


Figure 1. Diagram of the single base extension reaction. The incorporation of a complimentary ddNTP at the 3' end of the SBE primer is shown.

Different oligonucleotide lengths are used to assay each SNP and various fluorescent tags are used to display the interrogated nucleotide present in each allele. The intensity of the fluorescent products provides a measure of the amount of the DNA from each species present in the original sample. Fluorescent products are separated by size and detected by color using capillary electrophoresis instruments commonly available in today's crime laboratories.

Assays combining PCR and SBE can be very sensitive and support a high level of multiplexing (Inagaki et al., 2004; Onofri et al., 2006; Sanchez et al. 2003; Sanchez et al. 2005). Simultaneous detection of 35 human Y chromosome SNPs with as little as 100 pg of DNA (Sanchez et al. 2003; Sanchez et al. 2005) and single reaction typing of 37 Y chromosome SNPs using product from 6 PCR multiplexes (Onofri et al., 2006) have been achieved. Inagaki et al. (2004) typed 38 autosomal SNPs and one sex chromosome SNP in a multiplex SBE assay. Our approach will target simultaneous analysis of approximately 30 discriminating sites providing enormous advantage over the use of RT-PCR or immunologically based methods that cannot approach this level of multiplex analysis.

Bode will employ taxon-defining SNPs in three separate mtDNA genes. The cytochrome c oxidase I (COXI), cytochrome b (*Cytb*), and NADH dehydrogenase 4 (ND4) loci were chosen because literature review and our preliminary results indicate that an assay combining the canonical SNPs found within them will have the power to resolve differences between major clades, as well as between closely related taxa.

One of the advantages of using SBE to display the SNP polymorphisms is the ability to build high level multiplex sets for SNP interrogation. Since three separate genes are being evaluated, at least three PCR products will be designed using evolutionarily conserved sequences surrounding the variable regions containing SNPs. Depending on the frequency, lengths, and

location of conserved regions and the resulting amplicon lengths, more PCR sets may be required. (See figure, below.) Multiplex PCR development is a strength of the investigators so we anticipate we will be able, in the end, to design a single multiplex, or possibly two, to amplify all segments of interest. We will proceed as quickly as possible to multiplex PCR processes as this is the goal of the work and new challenges are often encountered at this stage negating "apparent success" with individual PCR products.

To prepare the PCR products for SBE, the PCR primers will be removed using the commercially available ExoSap-IT® PCR clean-up product. Each PCR product will contain SNP target sites that define variability across taxa. We will design SBE primers, as described above, of varying lengths that terminate one base proximal to each polymorphic SNP site. Size adapters ("tails") of non-hybridizing poly-T sequence will be added to the 5'-terminus of SBE primers to achieve substantial length variability without compromising hybridization conditions. Each SBE primer will be extended using AmpliTaq FS Polymerase in combination with two or more dideoxynucleotides each containing a respective fluorescent label to determine the nature of the respective downstream SNP. Simultaneous extension from many primer sets can be performed in one reaction simultaneously.

It is important to note that the same PCR multiplex product can be used in multiple SBE multiplex assays to increase the complexity of the interrogation. This is a key component of the design that not only increases the resolving power of our approach, but also streamlines the addition of future target sets that can employ the same PCR products in combination with new SBE interrogation sets to be designed in the future, thus expanding the menu of species targets

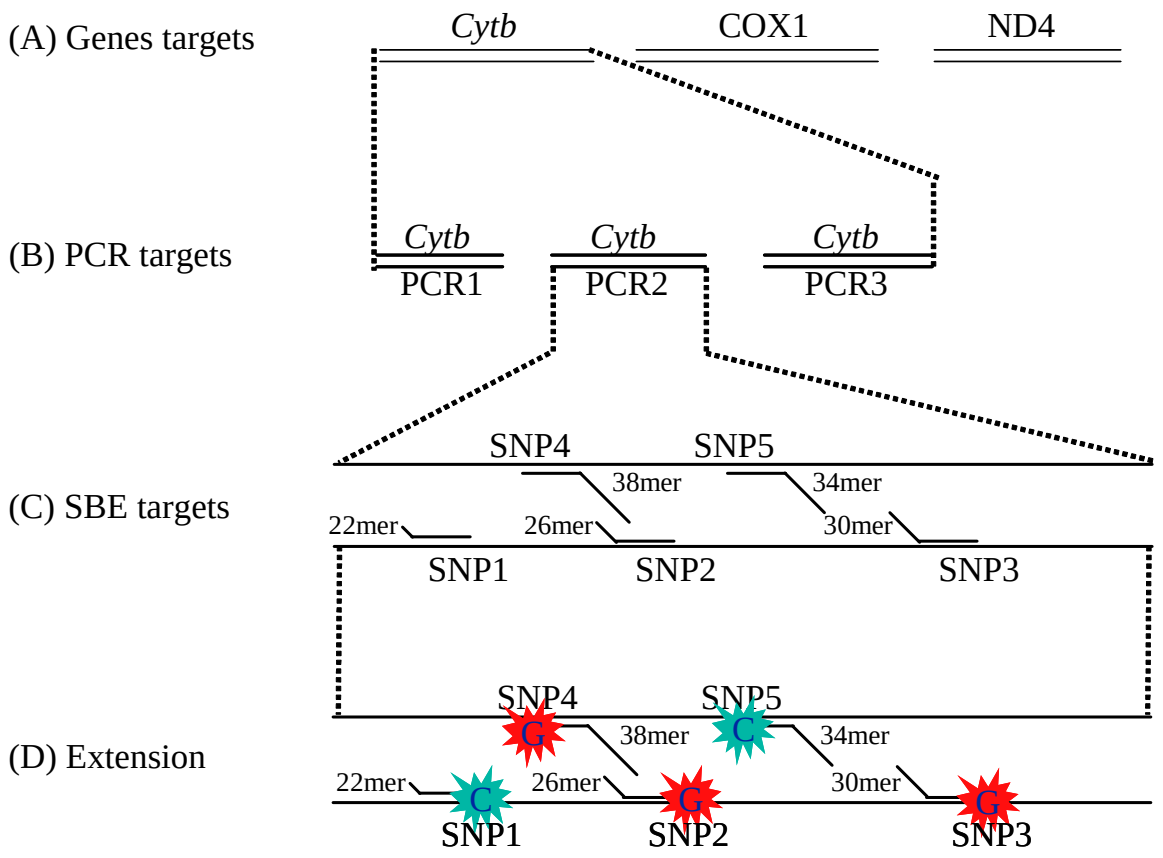


Figure 2. Assay components and structure. (A) Three target genes, *CytB*, COX1, and ND4, will be employed in the assay. (B) Each gene may contain multiple PCR segments that will be amplified using conserved sequences selected as PCR primers, shown here for the *Cytb* gene. (C) Within each PCR amplicon, one or more SNP sites will be interrogated as SBE targets. SBE primers will contain a specific hybridizing component that terminates adjacent to the variable SNP site of interest. Each SBE primer will also contain a non-hybridizing segment that is included to add length to the primer so that SBE products will be of different lengths depending on the site being evaluated. Note that primers can be used to interrogate either strand of the same amplicon. (D) Extension of the SBE primer using *Taq* FS Polymerase and fluorescently labeled ddNTPs will result in incorporation of fluorescent tags depending on the presence of a particular base at the extension site. The size of these products will be determined following separation via capillary electrophoresis and the base content at each site by the color of the product.

Assay Evaluation, Interpretation, and Control DNA

Assay compatibility with capillary electrophoresis instruments offers great advantage over microarray approaches that would require creation of a new nation-wide instrument infrastructure and associated complex training programs. SBE products will be separated by capillary electrophoresis using a 3100 or 3130xl Genetic Analyzer. The SBE set and the length of the SBE product will define each interrogated locus, and the fluorescent dye associated with that set and product length will indicate the base (i.e., the allele) present in the test sample.

The sum of all these SNP alleles can be displayed as a single string of interrogated bases (or several such strings) as shown in the figure, below. Sample content can be displayed as deviations from a consensus sequence in similar fashion to the use of the Anderson sequence with mitochondrial sequence description today (Anderson et al., 1981; Andrews et al., 1999). Thus individual animals comprising a particular group will generate a unique or small number of patterned sequences at these strategic sites.

This approach to representation of the results simplifies the interpretation of potential mixtures as mixed samples will contain the sum of the characteristic alleles present in the each of the separate sample taxa. This capability is an important advantage of our approach that would be difficult to achieve using other approaches. Of course, samples could be cloned and separated to permit sequencing, but such a protocol is time consuming, cumbersome, and requires sequencing from exponentially more colonies as more individuals are added to a mixture.

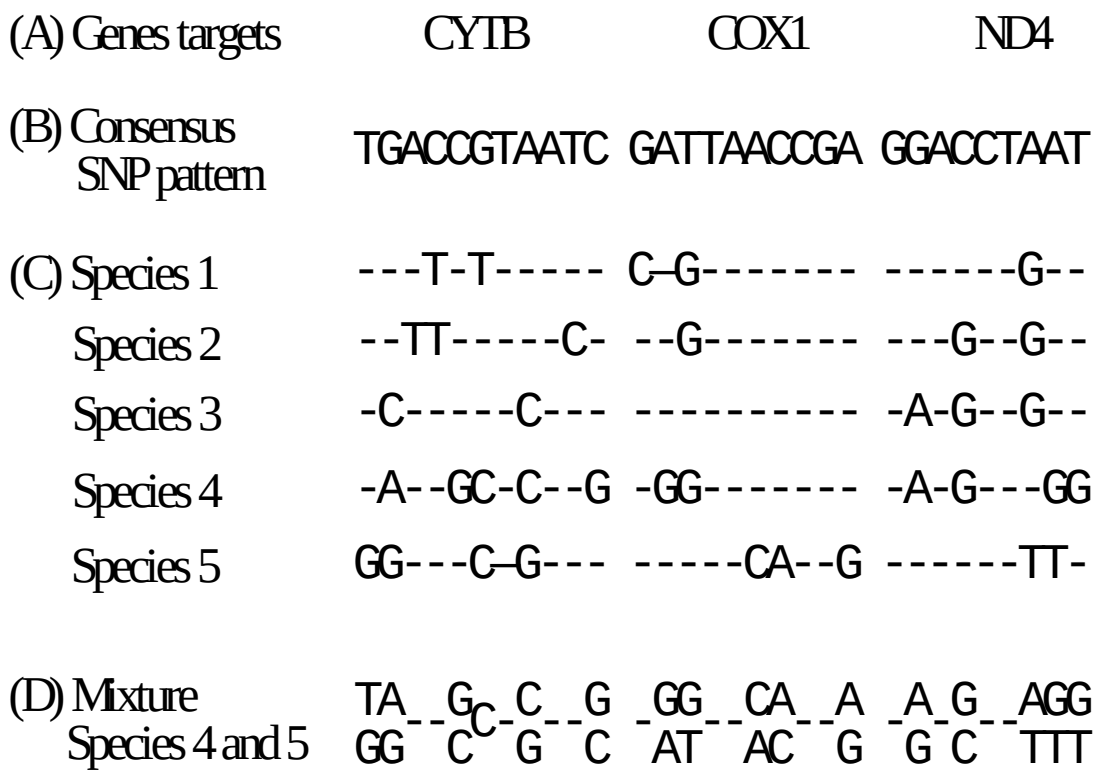


Figure 3. Hypothetical version of comparative sequence analysis and assay interpretation. (A,B) This illustration demonstrates a consensus sequence across species for the entire span of three genes. In this example, the *CytB*, COX1, and ND4 genes contain 11, 10, and 9 SNP sites of interest, respectively. (C) Individual specimens will contain particular alleles at each SNP site of interest. Sequences that are identical to the consensus can be shown as dashes, while those that vary from it can be defined by the base of the allele. Each different species will have a different sequence pattern or group of sequence patterns that characterize it. (D) A mixture of two DNAs will reveal some regions of commonality and other regions that show two different bases at one SNP site. In this case, the mixture contains Species 4 and Species 5 components. While the example shown here is more complicated than typically expected, it should be possible to decipher the content of a mixture from pattern like these. This should be especially simple if one of the components is known, e.g. human.

Creation of a molecular signature based on variation in three separate genes amplified in three separate reactions serves as an internal confirmation control. Anomalous or disparate results across gene panels should be detected immediately. During development, we will test the 10 or more sample DNAs representing each taxon and will combine samples across taxa to

demonstrate appropriate performance of the assay. In addition, we will endeavor to create positive control DNA to confirm both PCR and SBE performance during each individual assay. That is, DNA representatives of each taxon will be mixed together to create a "zoo DNA control" (or more correctly "zoo plus human DNA control").

Overview of Specific Goals

The specific goals of this project over the course of completion are as follows:

1.) Compile a data matrix of sequence from ten representative individuals across ten animal taxa native to North America commonly encountered at crime scenes for the entire sequence of three mitochondrial genes (COI, ND4, and Cytb). The species are to include ubiquitous household pets, livestock, yard animals, and some vermin.

2). Identify taxon specific canonical SNPs for use in identifying components of a single contributor sample or from a sample containing mixed contributors.

3). Develop single base extension markers (SBE) for use in an ABI Prism® SNaPshot™ based assay for each of the taxa specific canonical SNPs and "universal" polymerase chain reaction (PCR) primers across the three mitochondrial genes encompassing the SNP sites.

4). Identify best-fit reaction conditions for each amplification primer separately, check each primer against all taxa to ensure non-crossreactivity.

5). Multiplex the PCR primers into one initial reaction.

6). Design mini-multiplexes with SBE primer sets and optimize the individual reactions.

7). Develop conditions for optimum reaction efficiency with a positive "zoo" control.

Summary of Progress on Goals:

1) Filling out the data matrix using GenBank and in-house sequenced samples

In our proposal, we stated that we would use sequences from ten unrelated individuals from across ten distinct, forensically relevant taxa to create our assay. In the original proposal, the initial PI stated that the sequences would mostly come from previously sequenced samples found in NCBI's GenBank DNA database. Upon examining the database, it became clear that a good portion of the mitochondrial sequence we needed to complete our task wasn't to be found. Several samples important to the task such as white-tailed deer and domestic cats were conspicuously absent from the database. It was necessary to obtain biological specimens of these missing taxa, so that sequences could be generated to fill in the gaps in our data matrix. Sequenced specimens were obtained from in-house holdings here at Bode as well as from the forensic laboratory of the U.S. Department of Fish and Wildlife Services in Ashland Oregon. In total approximately 100,000 base pairs of DNA were sampled and sequenced to complete the data matrix.

2) Identifying taxon specific canonical SNPs

In order to determine SNP location through phylogenetic inference from species with such divergent evolutionary histories, it is first necessary to align the multiple sequences for each gene. Sequence alignments for this project were performed using ClustalX software. ClustalX creates .aln files of the sequence in an aligned format. The aligned files were then imported into PAUP* (Phylogenetic Analysis Using Parsimony) v4.0b software for phylogenetic analysis. Each of the genes was analyzed using a neighbor joining algorithm, then bootstrapped for 10,000 replicates. The canonical SNPs were mapped to each breakpoint using the "define trees" option.

The resultant panel of SNPs was compiled into spreadsheets and examined by eye in the alignments for uniqueness. Sites determined to be unique to a certain taxa were designated as "canonical" SNPs. Appendix A contains the gene of origin, location, base change, and corresponding SBE primer for each of the SNPs initially identified through phylogenetic analysis.

3) Develop both SBE and their corresponding PCR primers.

A subset of the best performing SNPs were chosen, and primers were designed just upstream (5') from the SNP. Because the SNP primers are to be taxon specific they had to work within their given taxa, but it was unnecessary to ensure "universal" utility. Therefore in all cases the SBE primers exactly match the preceding eighteen to twenty nucleotides of sequence preceding the SNP. This process was then repeated for the complimentary, or reverse, strand which can also be exploited. It simply gives genotypes as the complimentary SNP to that on the forward strand. Thus, two possible primers exist for each canonical SNP. This increases the likelihood of successfully exploiting these markers. .

Once the canonical SNPs were located with respect to overall gene length, areas of greater SNP concentration across all taxa and were identified and used to develop degenerate PCR primers to amplify these regions. The development of a large number of testable primers across the entirety of each of the three gene regions proved difficult. In addition to the primers designed in-house, "universal" Cytb and COI primers from literature were also included in the amplification tests as it was determined that the regions amplified by these "barcoding" primers encompassed a significant number of our SNP sites.

4) Optimize primer conditions and test for cross-reactivity

After the PCR primers were developed, the optimum reaction conditions for each primer pair were determined separately. Several factors including annealing temperature, annealing time, and ramping speed were examined until the conditions allowing each primer to amplify across the widest number of taxa/individuals were identified. Using the number of taxa amplified by a primer set and number of SNP sites contained within as selection criteria, a subset of four of the original PCR primers were chosen for further testing. The four chosen primers consist of one from the COI gene, one from ND4, and two from CytB.

Once a panel of PCR primers for the assay was chosen, each of the SBE primers was tested in multiple iterations against members of their corresponding taxa to ensure proper functioning individually. For each PCR amplicon, both the forward and reverse strand SBE primers were tested. Reaction temperatures and cycle durations were optimized during this testing to ensure efficiency and repeatability. At this point it became obvious that the use of degenerate "universal" primers would have an effect on the background noise of the electrophoretic visualization. A number of variables were examined, however no specific point of the assay was determined to be responsible for the background, and it simply appears to be brought about by the low PCR annealing temperatures inherent to this process.

After a first pass scoring to determine which SBE primers were the most efficient, the same primers were then tested against each of the other taxa in order to ensure non-cross reactivity. These cross-reactivity tests were performed to enable downstream multiplexing to proceed more smoothly. Any primers producing false positive peaks were excluded from use in the final assay. This was surprisingly effective, as the number of primers for each taxon was decreased significantly. Also during this time, it was determined that the Mustelidae taxa, which

included mink, wolverines, badgers, and otters is too diverse to be included in the final product. Through this testing we have developed a panel of eighteen SBE primers encompassing the nine taxa that will enable double coverage for each taxon. In other words, each taxon will have two SBE primers in the final assay to add confidence.

5) PCR primer multiplexing

This assay is intended to be an initial screening tool to enable a Forensic lab worker to quickly and efficiently determine the species of an animal. Because further testing may be necessary it is our goal that the assay be as sensitive as possible to ensure a minimal loss of sample. The overall sensitivity of the assay depends on the efficiency of the initial PCR multiplex. When adding multiple primer sets to one PCR reaction, competition is introduced, whereby each primer set competes with the others for DNA template to anneal to and reaction components. To test for competition, initially, stock of equimolar ratios of each primer at a ten micromolar concentration was used. Individuals from each taxon were amplified with the multiplexed PCR stock and then reacted with the corresponding SBE primers. Each SBE reaction was performed with the same concentration of SBE primer for uniformity. Yield gel band strengths along with the relative fluorescence units (RFUs) from each of the subsequent SBE electropherograms were used to determine the relative amplification efficiency of the multiplex.

The amplification cycling parameters finally settled on are a modification of a STR-type multiplex reaction with lower annealing temperatures and longer annealing time. Once each SBE was scored as present in the multiplex reaction from its corresponding taxa, the reaction conditions were considered successful, as SBE primer concentrations can be linearly altered to

modify signal strength. At the present, we have shown proper amplification from starting templates as low as .1ng of DNA contained in two microliters of extract, however in general reproducibility is still inconsistent with respect to the initial amplification reaction. For example from an input of 1ng of feline DNA, the RFU levels for SNPs 23R and 14R consistently appear in the 3 thousand to 4 thousand range. In contrast, the Gallus SNP 51F with a 1ng starting concentration of template will randomly produce RFU's anywhere between 300 and 3k RFU's.

6) Design and optimize SBE mini-multiplexes

In order to ensure that all eighteen SNP sites could be analyzed using only one PCR reaction, it is necessary to collect several SBE primers into multiplex panels. The multiplexing of primers into panels is based upon the exploitation of differences in primer length and fluorophore color. The SNaPshot^{TM} Multiplex Kit employs a different fluorophores, one for each nucleotide type, allowing the exploitation of color. In order to differentiate primers by size, for our assay, poly-T nucleotide tails were added to the 3' end of primers. The initial panel strategy consisted of three mini-multiplexes grouped based on the genes which the SBE primers resided in. Each mini-multiplex contained between five and eight SBE primers with no-overlap. Initially, all primers were added to the multiplex mixture in equimolar concentrations. This initial strategy was thought to be less effective due to the competition for common reaction resources and a second set of panels was designed with the opposite goal in mind, to reduce the number of primers from the same gene in each mini-set. This decreases the amount of competition for each PCR amplicon template in the SNaPshot^{TM} reaction. By having less SBE primers fighting for annealing space on each amplicon, we hypothesize that there will be faster, more efficient binding during the PCR annealing phase. This set of three SBE mini multiplexes

was used as the finalized assay for proof of concept. At this point, the SBE primers began to work in conjunction; however we were not able to get all of the peak heights even. There is simply too much difference in amplification rates for linear adjustment of SBE primer amounts to compensate for. Also the high background and spurious artifact occurrence is still an issue with the assay, and we are adjusting the procedure to attempt to decrease this noise.

7) Develop a "zoo" control

In order to ensure proper functioning of any PCR based assay it is necessary to implement experimental controls to monitor the effectiveness of both the assay itself as well as the technique of the scientist performing the assay. Proper functioning controls are vital if an assay is to be validated for use in any forensic laboratory and subsequently admitted into evidence in criminal procedure. The negative control for this assay consists of the entire PCR reagent mixture subjected to a round of PCR and SBE in the absence of DNA template. Positive peaks upon visualization are indicative of outside contamination in one or more of the reagents requiring the scientist to begin a new assay with fresh reagents. Proper quality control procedures on all reagents being brought into or created in the laboratory should identify contaminated items before use.

To ensure proper assay functioning it is also necessary to run a positive control ensuring that all components of both the PCR and SBE reactions are performing efficiently. Because of the nature of this assay, normal forensic PCR positive controls such as 9947A or 9948 human DNA are useless. In order to ensure that the PCR primers amplified successfully, a simple 2% agarose yield gel can be run and the bands can be scored at their corresponding lengths (Cytb1 – 350bp; Cytb2 – 405bp; ND4 – 418bp; COI – 523bp). However the absence of any one of the

four bands from a gel on its own is not a confirmation that the SBE reaction will not work. Therefore unless there is no visible amplified product, we recommend going forward with the SBE reaction. To provide a positive control for the SBE reaction we have attempted to develop a "zoo control" consisting of an amalgamation of DNA from each of the nine taxa identified in the assay. This control DNA is carried down stream through the experiment alongside the unknown biological samples. The "zoo control" requires laboratories to maintain fresh animal DNA from nine different species for the purposes of performing this assay. Also necessary is the extraction and Quant-iT^{TM} PicoGreen quantitation of DNA from the nine different animals. This requirement will most definitely be prohibitive for some laboratories, as some samples such as Black bears are more difficult to obtain. The methods by which large amounts of the control DNA would be obtained are outside of the scope of this project and the capabilities of Bode, and are not explicitly addressed in this report.

Executive Summary Conclusion and Future Directions

The experiments performed under this grant have developed the conceptual basis and proof of concept for a multiplexed assay for the identification of contributor species to forensic biological samples. Often in cases where non-human DNA presence in a sample is suspected, the method of species determination is performed through the sequencing of a small portion of the Cytochrome b mitochondrial gene, where the generated sequence is then queried against the NCBI Blastn database to determine species of origin. This method is quite functional; however it has several inherent flaws, including the inability of Sanger sequencing chemistry to analyze mixtures, as multiple nucleotides at one position are inseparable and will confound analyses. Also the BLAST similarity score, though powerful, is only a similarity score and is bound by the

species contained in the GenBank database. Often, as was the case when we began searching for sequences to fill out our data matrix, seemingly ubiquitous species are absent. In these cases, a similarity score would be unable to determine the species origin.

The PCR-SBE-SNP assay and corresponding primer panels developed under this contract enable the identification of species of origin from either single or multiple contributors. The overall assay format is very similar to STR throughput and ultimately should fit in seamlessly with the casework analyst's arsenal of tests. The assay requires as little as one hundred picograms of DNA template per species in a two microliter volume of extract to produce an accurate result (the amount of template required increases with the number of species in a mixture). We have been unable to fine tune the SBE multiplexes to produce equal electropherogram RFU peak heights, as amplification bias seems to be random and changing SBE primer amounts decreases the sensitivity of the test. Experiments attempting to determine the origin of spurious artifacts and elevated background noise in the resultant electropherograms to date have provided no easy answers for these problems. Unfortunately, at this point it appears as though the SNaPshot™ assay is not meant to be reacted against amplicons generated through the use of degenerate primers under extreme PCR conditions and anomalies will be present in any version of this assay. It is worth considering that the continuing presence of artifacts may be due to limitations inherent to the use of the SNaPshot™ kit. Other methodologies for the identification of SNPs should be considered for use with these SNP sites, including but not limited to Roche 454 Life Sciences pyrosequencing, Illumina Solexa, and ABI SOLiD. The platform issue lies with the ability of each of the techniques to resolve mixtures of SNPs.

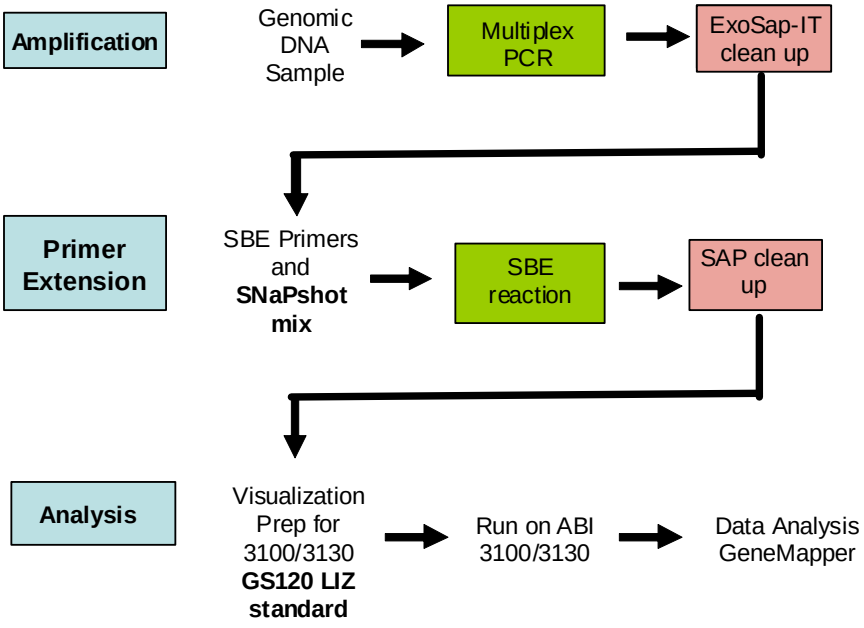


Figure 5. Work flow diagram of the PCR-SBE-SNP assay showing each step of the proccess. Thermal cycling (green) and clean up (red) reactions are denoted as they take up the majority of the proccess temporally.

Experimental Details and Main Findings

Sample Gathering and Data Matrix Generation

Bode's plan for assay development included significant utilization of bioinformatic tools. One of the tools at our disposal is the GenBank genetic database maintained by the National Center for Biotechnology Information (NCBI). GenBank is a public repository for sequence and protein data representing the genetic diversity found across diverse organisms from unicellular bacteria all the way through the Human genome. As of August 2009 there were 108,431,692 sequence records in the traditional GenBank Division. In the original proposal, it was stated that we would collect sequences from ten unrelated individuals from across ten distinct, forensically relevant taxa to create a data matrix. Upon examining the database, it was evident that mitochondrial sequence data for some animal taxa commonly found at crime scenes, such as domestic cats and the ubiquitous white-tailed deer, were virtually absent from the database. In

several cases, the database entries for certain animals consisted of “DNA barcodes”, which are small gene fragments (200-400bp on average) used for taxonomic analysis. Due to the nature of the SBE assay, we were confined to using only the portions of the genes which “universally” amplify using a degenerate PCR reaction. By utilizing the entire gene regions, more areas amenable to the development of PCR primers can be identified and thus increasing the chance of success by incorporating the most canonical SNPs possible into the final assay.

The initial sequencing portion of the project began with the procurement and sampling of biological material from several individuals in each of the taxa. Several specimens were available in research and development department at Bode. Requests were made of the Museum of Natural History at Louisiana State University and the Department of Fish and Wildlife Services, for the remainder of specimens required for completion of the task. Letters of support were received from both institutions granting access to specimens. The forensic laboratory of the Department of Fish and Wildlife Services was able to provide all of the specimens of interest to us.

The samples received from the Department of Fish and Wildlife Services for testing consisted of 1.5mL micro-centrifuge tubes of either tissue or blood from the animal of interest. The samples on-hand at Bode consisted of frozen tissue and hair from the various animals in 1.5mL micro-centrifuge tubes. All of the samples were extracted with the Qiagen QIAamp® DNA Micro Kit and eluted into 50ul of H_20. The entirety of each of the samples was consumed during nucleic acid extraction. At present, we do not have a real-time method for quantification of mitochondrial DNA outside of humans; therefore a total DNA quant was performed. Each of the samples were quantified for total DNA content using the fluorescent double strand intercalating Quant-iT™ PicoGreen dsDNA Reagent on a CytoFluor® Series 4000 multi well

plate reader. This reagent intercalates into any and all double stranded DNA, thus the quant provides the total DNA (nuclear and mitochondrial) contained in a sample. From these assay results we were able to gauge a rough extrapolation of mitochondrial DNA content from the total DNA content.

Table 1. A list of specimens sequences for this project. Sample numbers from the Department of Fish and Wildlife Services are listed as the original sample number.

Taxa	Genus	species	Original Sample #	Bode Sample #
Cervidae	Odocoileus	hemionus	(in-house)	Ce1
	Cervus	canadiensis	(in-house)	Ce2
	Alces	Alces	(in-house)	Ce3
	Odocoileus	virginianus (WI)	(in-house)	Ce4
	Odocoileus	virginianus (MI)	(in-house)	Ce5
	Odocoileus	virginianus (MI)	(in-house)	Ce6
	Odocoileus	virginianus (ND)	(in-house)	Ce7
	Odocoileus	virginianus (ND)	(in-house)	Ce8
Mustelidae	Neovison	vison	(in-house)	M1
	Nevoson	vison	D40422	M2
	Nevoson	vison	D40414	M3
	Gulo	gulo	G20765	M4
	Gulo	gulo	C11371	M5
	Enhydra	lutris	J10914	M6
Ursus	Ursus	americanus	K10757	U1
	Ursus	arctos	H21051	U2
Felis	Catus	domesticus	(in-house)	F1
	Catus	domesticus	(in-house)	F2
	Catus	domesticus	(in-house)	F3
	Catus	domesticus	(in-house)	F4
	Catus	domesticus	(in-house)	F5
	Catus	domesticus	(in-house)	F6
	Catus	domesticus	(in-house)	F7
	Catus	domesticus	(in-house)	F8

In order to amplify the three genes from each of the animals not represented in the database, available sequences from their nearest neighbors evolutionarily were compiled and used to develop degenerate primers. All primer design for this project was performed with

Primer3 version 0.4.0. This software allows the user to upload a region of DNA sequence and analyze the placement of PCR primers or probes based on general primer picking conditions including optimal primer length, melting temperature, %GC content, and self complimentarity. It was necessary to generate three or more primer pairs for each gene, from each taxon to sequence the approximately 1200 to 1500 bases making up each of the three genes. In some difficult cases, one micro-liter of bovine serum albumin (BSA) was added to non-responsive reactions to generate product.

All of the samples sequenced for the matrix, were amplified using an annealing temperature of 45°C for 40x cycles using an ABI 9700 GeneAmp® PCR system. The PCR mixture components and volumes used are specified in Table 2a. (below). Each sample was then cleaned up through use of ExoSap-IT® reagent and used as template in a sequencing reaction. Sequencing was carried out using ABI BigDye® Terminator v1.1 Cycle Sequencing Kit with the reagent volumes listed in Table 2b. The resultant sequences were run through Edge Bio Performa DTR gel filtration cartridges to remove all unincorporated dye terminators and visualized on an ABI PRISM® 3100 Genetic Analyzer. Sequence analysis was performed using Sequencher v4.7 software.

Table 2. Reagent component volumes for both a) amplification and b) sequencing reactions.

a)

Amplification Component	Volume (ul)
dH_20	11.91
Forward Primer	0.75
Reverse Primer	0.75
Gold Star Buffer (Promega)	1.5
Amplitaq Gold (Promega)	0.09
DNA Template	2
	17

b)

Sequencing Component	Volume (ul)
Big Dye v1.1	1
Better Buffer	5
Sequencing Primer (3.3mM)	0.75
dH20	6.25
DNA Template	2
	15

Due to the use of degenerate primers for sequencing, special care was taken in the analysis of the sequence obtained from the samples. Because of the reduced annealing temperatures used it was imperative to ensure that no non-specific amplifications had occurred. All sequence electropherograms were analyzed by eye and checked for any irregularities including bad sequence and heteroplasmies. The resultant sequence was then checked against the NCBI database and the nearest neighbors were identified to ensure that the samples were the proper gene and taxa expected. In total, with the in-house and FWS samples finished, just over 100,000 base pairs across twenty-four samples were sequenced to complete the data set. A list comprising taxa and species is included in Table 1. All of the sequences were catalogued and translated into both FASTA and #NEXUS file formats for use in downstream bioinformatic applications.

Bioinformatic Analysis and Canonical SNP Identification

The sequences obtained from GenBank were added to the in-house generated sequences and compiled into three FASTA formatted files, one for each gene region. These three files comprise the data matrix that we will be using to determine the canonical SNPs for each taxon. The files were individually imported into ClustalX v2.0.10 in order to perform the complete sequence alignments. The figure below (Fig. 4.) shows the alignment of a region of cytochrome B with each of the base pair types color-coded to enable the user to pick up alignment patterns visually. This was done independently for each sequence. Gap penalties were examined and weighted to determine which settings produced the most accurate alignments.

Figure 6. ClustalX alignment of sequence showing color-coded base pairs exhibiting regions of similarity and dissimilarity visually.

Once sequence alignment gaps were set, nearest neighbor joining (NJ) trees of each of the genes were created. The neighbor joining algorithm works off of a minimum evolution framework, wherein trees showing the minimum branch length between taxa are preferred and scored accordingly. Because we are building a tree between such highly divergent taxa, this method should adequately resolve relationships in a timely manner. In the resultant trees, each of the taxa coalesces to one node. This strict bifurcation is important to the discovery of canonical SNPs, as these nodes will set the break points responsible for taxonomic inclusion. There are no anomalies in the tree, meaning that, none of the genes have any cryptic species more closely related to other taxa than to their own. Lastly, the aligned files were used to create consensus by pleurality consensus sequences. That is, using a majority rule, a consensus sequence was created for across each of the three sequences to use as a basal reference sequence.

This will allow us to relate the SNPs identified in the assay to the “ancestral” consensus base at a position.

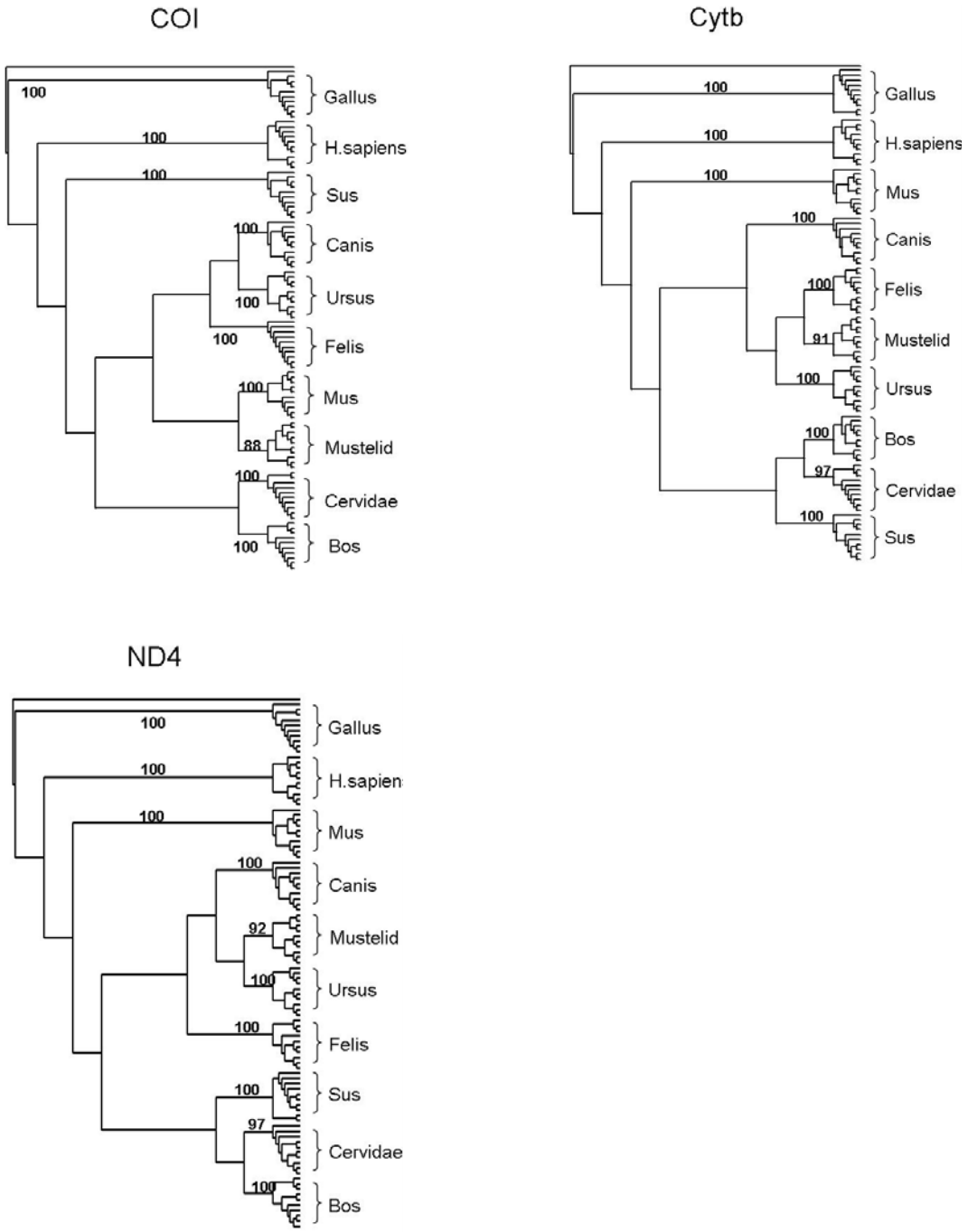


Figure 7. Non-parametrically bootstrapped neighbor joining trees, for each of the three genes, across all included taxa, showing proper taxa recovery and separation. The bootstrap values from 10,000 replicates are indicated above the corresponding branch lengths.

The aligned files (.aln) were then exported into PAUP* 4.0b phylogenetic software. Each of the three gene files were analyzed separately using the neighbor joining algorithm, then bootstrapped for 10,000 replicates. The resultant consensus trees, along with the resultant bootstrap values for each of the taxon break points are shown in Figure 7. Once the individual gene tree topologies were recovered and bootstrapped, the canonical SNPs could be mapped to each breakpoint. The PAUP software allows the user to break down each tree and identify the apomorphies (a feature or character unique to a clade and its descendants; for our purposes - a canonical SNP) responsible for resolving individual branches. The branch, character number, change, number of evolutionary steps, and consistency index (CI) are all listed for each apomorphy.

a. b.

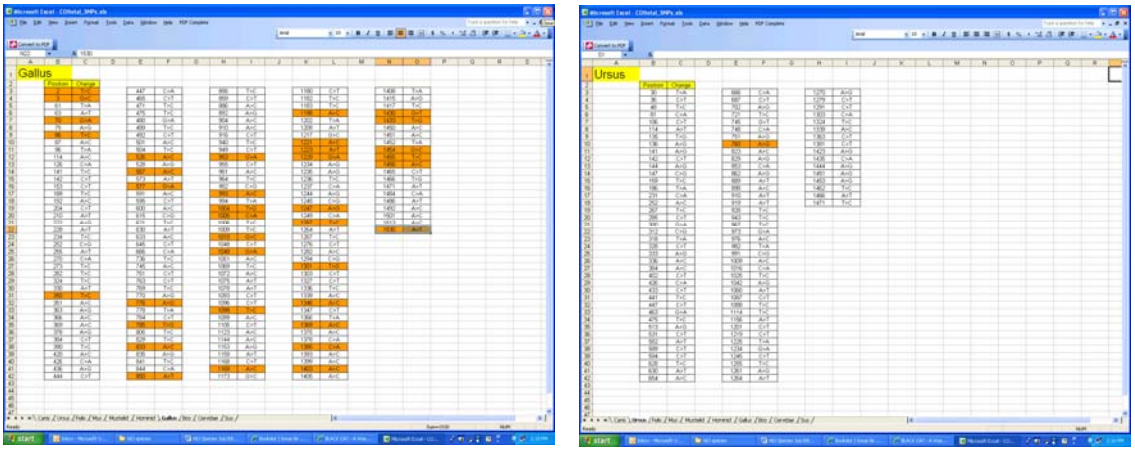


Figure 8. Screen captures of the (a) Gallus and (b) Ursus taxa SNP excel spreadsheets. SNP sites with a CI of 1.0 are highlighted in orange. Clearly Gallus possesses more useful SNPs.

Because the taxa in this analysis are so distantly related, for any one of the three trees there exist several thousands of apomorphies. Since each of the SNPs had to be screened by eye to determine uniqueness, the consistency index was used as a first screen to decrease the total number of SNP locations examined. This consistency index represents a measure of "goodness

of fit" for the SNP in resolving the related branch. Only those SNPs with a CI rating of 1.0 were screened for inclusion in the SBE reaction. Excel spreadsheets were compiled listing each SNP site from each of the three genes across all ten taxa. A great deal of variation in number of useful SNPs can be seen between taxa. The SNP data comprises thirty Excel spreadsheets which are available for viewing. It should be noted that further work utilizing this method will certainly seek to increase the number of taxa identified. We examined nine of the hundreds of taxa that invariably would be included in an "optimal" species test. The eventual addition of more species to this analysis will not invalidate these results, but rather, will expand the understanding of the resolving power of each of the SNPs. Future work will be able to use our assay as an initial screen and then focus on identifying the species within each taxon in a more detailed fashion.

SBE and PCR primer development

A subset of the best performing SNPs were chosen, and primers were designed just upstream (5') from the SNP. When designing SBE primers for use in an assay which utilizes liberal cycling parameters for "universal" amplification of source material, some aspect of normal mini-sequencing primer design had to be bent. For instance, it is prudent to design primers with an approximately equal percent GC content. This means that the percentage of the oligo-primer stretch that is comprised of guanine and cytosine bases should be similar to that of other primers if they are to be used in the same reaction. In development, we were forced to use primer pairs with GC contents varying by up to 20 percent. Unlike with the PCR primers, the SBE primers were only required to work within the diagnostic taxa, therefore primers could be designed without taking the consensus sequence into consideration. We were able to keep the primer lengths between 18 and 20 nucleotides long in order to ensure adequate specificity with

minimal annealing time. Also, DNA as a double stranded helix possesses an opposite strand of complimentary DNA to the forward or coding strand, originally used to identify the canonical SNPs. The complimentary nucleotide to the SNP position on the forward strand has the same power of identification as that of their counterpart. Therefore, this process of designing SBE primers 5' to 3' from SNPs of interest was repeated for the reverse strand. Thus, two independent primers exist for each canonical SNP, one forward and one reverse. A table of SBE primers, broken down by gene is included in Appendix A.

Once the position of individual canonical SNPs were located with respect to the genes themselves, focus shifted to areas of greater SNP concentration in order to attempt to develop degenerate PCR primers for these regions specifically. The primer developing program Primer3 v0.4.0 was then used to identify the regions in the consensus sequences that were most suitable to amplification. Primer3 lets the user define optimality criteria, thus reducing the majority of trial and error wet bench work through *in-silico* design. This region could then be examined in the alignment, and base locations where variation was seen were identified as degenerate (2 fold or 3 fold) and included in the primer design as such.

The development of a large number of testable primers encompassing each of the three gene regions proved difficult. For each gene only a handful of small regions appeared usable even when optimality criteria were loosened substantially. Of all of the genes, ND4 proved by far the most difficult to find primers for. This phenomenon is mirrored by the paucity of ND4 sequence contained in the NCBI database relative to the other genes used. This is due to the more rapidly evolving nature of this gene, as it exhibited by far the most diversity both within and between species. In addition to the primers designed in-house, the decision was made to include the "universal" Cytb and COI primers from literature in the amplification tests (Table 3).

Each of the primers was then tested on several individual samples from all of the included taxa. Both different species as well as multiple individuals from within single species were used to test each of the primers to ensure no binding mutations or significant amplification bias.

Table 3. List of PCR primers broken down by gene, showing the primer direction, sequence, and location of 5' starting base.

Gene	Primer Number	Direction	Sequence	5' Location
COI	1	Forward	ACAGCCCATGCATTYGTAAT	194
COI	1	Reverse	CCTCCAGCTGGGTCAAAGAAGGT	674
COI	2	Forward	TACTATCCCTACCAGTCCT	605
COI	2	Reverse	TAGTATAGCAGGAGATCATTT	1017
COI	3	Forward	GTAGGAGGCCTAACGGGAA	1048
COI	3	Reverse	TGATATGGTGGAGGGCATC	1509
COI	Universal	Forward	TCAACCAACCACAAAGATATTGG	25
COI	Universal	Reverse	TAAACTTCAGGGTGTCCGAAGAATCA	731
Cytb	1	Forward	AAAACCCACCCACTAATAA	16
Cytb	1	Reverse	GTAGGATTACTCCAATGTTTC	358
Cytb	2	Forward	CCGATACATACACGCAAACG	237
Cytb	2	Reverse	TTCGTGGAGGAATAGGAGGTGG	606
Cytb	3	Forward	CGCCTTCCACTTCATCCTG	537
Cytb	3	Reverse	AATAGGCATTGGCTGAGTGG	974
Cytb	4	Forward	CCATATCAAACCCGAATGAT	798
Cytb	4	Reverse	GTGGAGGCTGTTGCTTC	1234
Cytb	Universal	Forward	CCATCAAACATCTCAGCATGATGAAA	70
Cytb	Universal	Reverse	TCCTCAGAATGATATTTGMCCTCA	425
Cytb-bos	NA	Forward	TTAGGGGCCCTCTTGCTAAT	707
Cytb-bos	NA	Reverse	GGTGTTCGACTGGTTGTCCT	1017
ND4	1	Forward	AACATGACTATCAAAACC	41
ND4	1	Reverse	TGGTTTCCTCATCGGGTGATA	416
ND4	2	Forward	TAACCTTCACCGCCACAGA	323
ND4	2	Reverse	GCTTCTACATGGGCTTTTGG	668
ND4	3	Forward	CCAAAAGCCCATGTAGAAGCCC	649
ND4	3	Reverse	ATAGTTCGGCTGTGGATTCG	1025
ND4	4	Forward	CGAATCCACAGCCGAACTAT	1006
ND4	4	Reverse	GCTATGAGGGCATTTTCTCGTG	1313

A number of different PCR amplification parameters were examined during the initial primer screening process. An annealing temperature of 45°C, considerably lower than what is

normally used, was decided upon, as it allowed the primers to sit down across the widest range of taxa with minimal mis-priming. There seems to be a trade-off being observed between the PCR annealing temperature allowing for amplification across the widest range of taxa, and the level of background noise being seen in the subsequent SBE electropherograms. Next, the total number of PCR cycles was examined. Holding the number of cycles at 40 produced the best results with ample starting DNA. With lower copy number samples it was necessary to move the cycle number up to ensure proper amplification. Primer effectiveness was determined through the use of 2% agarose yield gels. The presence, intensity, and number of bands of correctly sized bands for each sample were then scored. If greater resolution is required, it may possible to use a flatbed, Hitachi style, scanner to more accurately size the amplicons, however that is beyond the scope of this research project. Table 4 shows the results of primer testing.

Table 4. A representation of the success/failure of the various universal PCR primers developed for this project and gathered from literature. Gray shading indicates an amplicon of the expected size was produced by the reaction, white indicates a negative response.

	Ursus	Mustelid	Cervidae	Canis	Bos	Felis	Sus	Gallus	Human	Mus
COI_1										
COI_2										
COI_3										
COI_Univ										
Cytb_1										
Cytb_2										
Cytb_3										
Cytb_4										
Cytb_Univ										
Cytb_B										
ND4_1										
ND4_2										
ND4_3										
ND4_4										

Both of the universal primers taken from the literature were successful in amplifying across the samples. In addition, in-house designed primers COI_1, Cytb_2, and ND4_3 each amplified at least nine of ten taxa respectively. The primers, COI_1 and universal COI, overlap

substantially and are for the most part interchangeable with regards to utility with respect to our SNP sites. None of the first round of primers that worked well across species included regions with a Bos (Cows) taxon canonical SNP. The length of sequence amplified with COI_3, included a large number Bos taxa specific SNP sites. Unfortunately, this primer functioned poorly across the assay and did not amplify in Bos *taurus* in any of our tests. A new cow specific Cytochrome B primer was developed (Cytb-B in Table 4.) which amplified well in Bos and across more than half of the other species. Thus this primer was added to the assay. Bode recognized the aspect of this addition, however at the point when this problem was observed; it was not feasible, time-wise, to explore the use of other mitochondrial genes.

SBE Primer Specificity and Cross Reactivity Testing

After PCR primers were developed, testing began for species specificity of our SBE primers. The first experiments were to determine if the developed SBE primers functioned properly within their dedicated taxa. Samples were amplified with one of the PCR primer sets using the 45°C annealing temperature cycling program for 40 cycles. All samples were cleaned via ExoSAP-IT® reaction and used as template in an SBE reaction using the ABI SNaPshot™ Multiplex Kit according to the manufacturer's standard recommendations (Table 5). Reactions were cleaned again using a shrimp alkaline phosphate (SAP) reaction. Size standard (LIZ 120) and HI-DI formamide were added to each sample and they were run on an ABI 3100 Genetic Analyzer using a 36cm capillary array. Sample files were exported into GeneMapper v3.2 and analyzed.

Table 5. SNaPshot™ reaction components and volumes for SBE reactions.

SNaPshot™ Component	Volume (ul)
SNaPshot™ kit	5
SBE Primer (0.2uM)	1
H20	1

PCR Product	3
	10

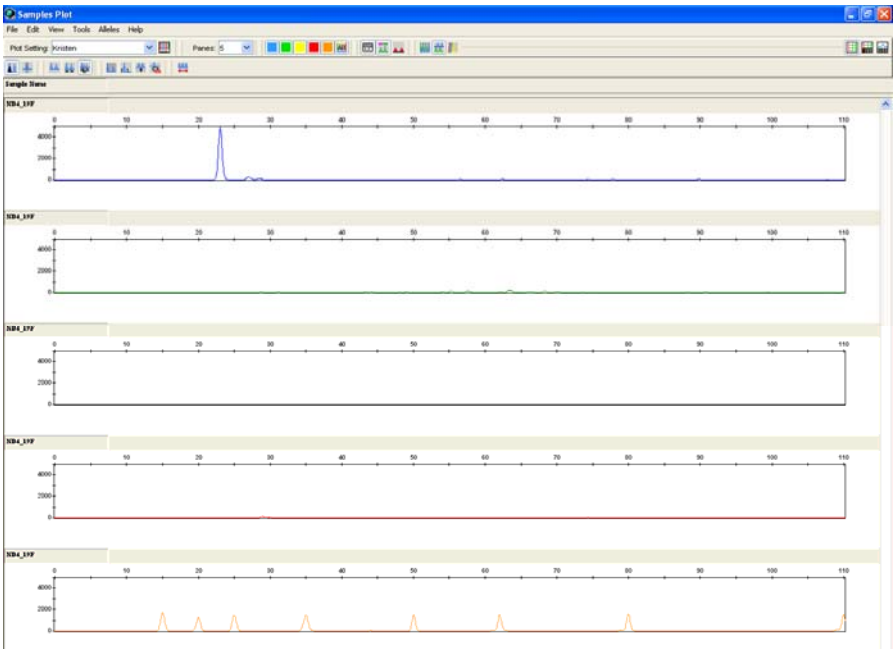


Figure 9. An electropherogram showing the Ursus 19F primer. The blue peak indicating the presence of a G at base 752 is clearly visible. (in SNaPshot™ A = green; C = black; G = blue; T = red) Primer 19F is 18bp long; however the positive peak appears as 23bp long.

A peak at the expected location was scored as a positive result and any other peaks within the examined range not considered pull-up or artifacts were recorded. An issue with the relative size of SNaPshot™ products became apparent; when visualized they appear slightly different sizes than they should based on a primer length plus one model. Thus for the introductory test, clear peaks above 150 RFU within six base pairs of the expected primer location were scored as a positive result. For each PCR amplicon, both the forward and reverse strand SBE primers were tested. The electropherograms of SBE primer results are numerous and will not be shown here

(however the printouts will be maintained on file at Bode). Figure 9 provides an example of a positive result.

Once the initial screen determined which SBE primers worked correctly and efficiently, the same primers were then tested against each of the other taxa in order to ensure that there was no cross reactivity. Each of the SNP sites were examined by eye in the gene alignments to ensure that they were unique to a particular taxa, so there would be no false positives for any of the markers. However, due to the use of degenerate primers and the extreme annealing conditions during PCR, Bode was aware of the very real possibility of mispriming or other spurious artifacts confounding analyses. Thus, these cross-reactivity tests were performed to enable downstream multiplexing to proceed more smoothly. Any primers exhibiting peaks of the expected color within reasonable distance of the expected SNP location were excluded from use in the final assay.

The cross-reactivity tests were surprisingly effective and the number of primers for each taxon was significantly decreased. During this portion of the assay development, it was determined that the Mustelidae taxa, would not be included in the final assay. The decision to exclude this taxon was based on the difficulty of getting primers to work properly both across and within the species comprising the taxa. The wide range of the Mustelidae family, which includes species as diverse as sea otters and wolverines proved too diverse for proper testing. The amount of genetic distance within the family was comparable to that between it and some of the other taxa, thus adding to the general erratic and irreproducible nature of its SBE markers. Finally, it was determined that the Mustelidae taxa contained the least useful species from a forensic standpoint, therefore we felt no overall loss from the exclusion.

Table 6. A list of the best performing SBE primers for each taxon. These primers will be used to develop the final working assay.

Taxa	Working SBE Primers
Mus	18F, 10R,
H.sapiens	55F, 56F
Bos	63F, 64F
Sus	62F, 26R
Ursus	19F, 13R
Gallus	51F, 6R
Cervidae	13F, 7R
Canis	3F, 2R
Felis	14R, 23R

Through several rounds of testing a panel of eighteen most robust SBE primers has been identified encompassing the nine taxa enabling double coverage for each taxon. In other words, each taxon will have two SBE primers in the final assay to add confidence. For certain taxa, such as Sus, it was not possible to find two independent SNP loci to use in the assay. In this case both forward and reverse strand SBE primers were used to confirm the presence of the taxa. This redundancy may not be used in the final assay, as a single SNP peak is sufficient to call the presence or absence of taxa; however both forward and reverse strands are being accounted for in the multiplexing portion of our experiments as though they will both be present.

PCR Primer Multiplexing

The sensitivity of this assay depends solely on the ability of the PCR reaction to simultaneously amplify minute amounts of mitochondrial DNA into enough template to facilitate reaction with eighteen different SBE primers. It was necessary to multiplex primers COI_1, Cytb_U, Cytb_B, and ND4_3 (from Table 4) into a single amplification. A great deal of time was spent on designing the initial primers; there were no indications of primer mis-pairing or primer self priming. In the first attempts at multiplexing, a modification of an STR multiplex was used as a base for the thermal cycle program conditions. Primer concentration was kept at 10uM for each primer in the set. The annealing temperature was kept at 45° C to maintain "universal" amplification. Figure 10 shows the results of the first multiplexing attempt.

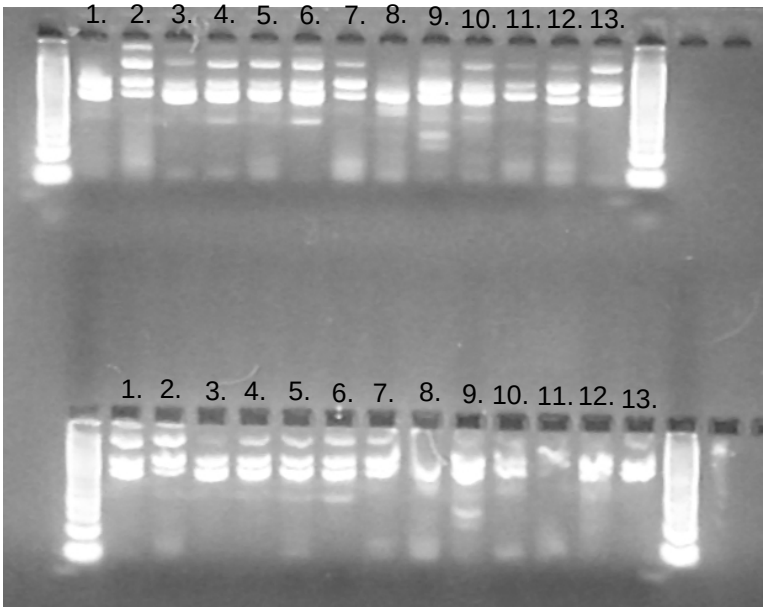


Figure 10. A 2% Agarose gel showing amplicon yields for different species from each of the ten taxa in duplicate, with 50bp ladder flanking each iteration. The top tow of samples used a 1ng template input, the bottom 0.5ng. 1. *Ursus americanus* (Black Bear) 2. *Ursus arctos* (Brown Bear) 3. *Bos taurus* (Domestic Cow) 4. *Odocoileus virginianus* (White-tailed Deer WI) 5. *Odocoileus virginianus* (White-tailed Deer MI) 6. *Odocoileus virginianus* (White-tailed Deer ND) 7. *Homo sapiens* 8. *Mus musculus* (House mouse) 9. *Felis catus* (Domestic cat) 10. *Canis lupus* (Gray Wolf) 11. *Canis lupus familiaris* (Domestic Dog) 12. *Gallus gallus domesticus* (Domestic Chicken) 13. Sus scrofa *domestica* (Domestic Pig).

Table 7. Finalized cycling parameters for a multiplexed PCR reaction.

# Cycles	Conditions
1x	95°C for 11 min, then:
	96°C for 1 min, then:
10x	94°C for 30 sec
	45°C for 30sec
	70°C for 45sec
30x	90°C for 30 sec
	60°C for 30 sec
	70°C for 45sec
1x	60°C for 30 min
	4°C soak

With only a few initial problems, the first pass at PCR multiplexing proved extremely fruitful. The reaction reagent concentrations were the same as those used for the individual amplifications (Table 2). It was necessary to adjust the amount of only one of the four primers, COI_1 to enable the SBE primers based on that particular amplicon to be reproduced efficiently in the subsequent reaction. Primer COI_1 was raised to 15uM concentration whereas the rest of the primers remain at 10uM concentration in the finalized protocol. The finalized thermal cycling profile is shown in Table 7.

SBE Primer Multiplexing

Once tested for efficacy in a single reaction, each of the individual SBE primers was incorporated into a theoretical multiplex scheme. The Original goal was to multiplex all eighteen primers into a single multiplex. However, to develop a working assay under the time constraints, it was determined appropriate to design the multiplex to function as a single entity as well as three smaller mini-multiplexes which would be easier to troubleshoot. To begin, each of the primers from Table 6 was broken down by SNP color and average RFU height during the

initial testing. In order to separate each of the primers temporally along a capillary electrophoresis run, poly-T (thymine) nucleotide tails were added to the 5' end of each SBE primer. These are naturally occurring nucleotides which don't react with most DNA, thus useful for primer spacing. Poly-A tails occur naturally on mRNA strands in mammalian cells, however, working with purified DNA it was anticipated that no interactions would be seen. Spacing order was determined by average RFU height, as it was thought that SBE primers exhibiting more robust RFU heights would be affected less by longer poly-T tails. The first group of spaced primers is located in Table 8 below.

Table 8a. Solo SBE primers grouped by gene of origin, showing the nucleotide position of the 3' end, SNaPshot™ Color, RFU height, and sequence of each.

Gene	Species	Primer	Position	Color	RFU	5'-3' Sequence
COI_1						
	Cervidae	13F	643	**red**	1881.5	CTATTAACAGACCGAAAT
	Canis	3F	606	**blue**	5267	ACTATCCCTGCCTGTACT
	Canis	2R	586	**green**	4376	TAATACAGGCAGGGATAGTA
	Gallus	6R	350	**blue**	567	TGCCGGCCCCAGCTTCTACG
	Cervidae	7R	643	**green**	252	CGAAAAAGGTTGTGTTTA
CytbU						
	Human	55F	137	**black**	1344.5	GCCTGATCCTCCAAATCA
	Human	56F	166	**red**	900	TTCCTAGCCATGCACTAC
	Sus	62F	366	**black**	438	AGTAGTCCTACTATTTAC
	Felis	23R	153	**black**	176	TCTGATGTGTAGTGTATGGC
	Sus	26R	366	**blue**	5586	ATGAAGGCTGTTGCTATAAC
	Gallus	51F	241	**green**	165	TACGGCTGACTCATCCCGG
ND4_3						
	Mus	18F	739	red	548.5	TACGGAATAATTCGCATC
	Ursus	19F	752	blue	526	GAATTACAACACTACTTG
	Mus	10R	718	red	1196.5	GATGCGAATTATTCCGTAAC
	Ursus	13R	764	black	340.75	TGAAGGGGTAAGCCATGAAG
	Felis	14R	759	blue	4357	GGTATGCCATTTGGTTCGTT
Bos (cytb)	Bos	841F	841	red	289	TTATTTGCATACGCAATC
	Bos	980F	980	black	2874	GCCAATGCCTATTCTGAG

Table 8b. Spaced SBE primers showing primer location and SNaPshot™ color, original primer length, original SBE product length, number of poly-T spacers added to the 5' end of each primer, theoretical new length, and empirical new length.

Gene	Primer	Color	Primer Length	Product Length	Spacers Added (T)	New Length	1st set real length
COI_1							
	13F	red	18	24.3	44	62	64
	3F	blue	18	24.3	36	54	54
	2R	green	20	27.1	37	57	56
	6R	blue	20	25.21	14	34	37
	7R	green	18	24.7	20	38	42
CytbU							
	55F	black	18	24.16	10	28	30
	56F	red	18	26.1	26	44	48
	62F	black	18	23.75	31	49	52
	23R	black	20	27.6	0	27	27
	26R	blue	20	24.09	44	64	67
	51F	green	19	29	0	29	29
ND4_3							
	18F	red	18	25.28	14	32	36
	19F	blue	18	23.32	0	23	23
	10R	red	20	27	41	61	60
	13R	black	20	26.3	16	36	40
	14R	blue	20	26.7	25	45	49
Bos (cytb)	841F	red	18	24.4	0	24	24
	980F	black	18	27.24	42	60	63

There is a phenomenon which occurs in SNaPshot™ primers where the length of the nucleotides from the primers themselves do not exactly correlate with the product length in the assay itself when viewed on an ABI 3100 CE machine. This phenomenon is even present in the positive control samples processed alongside unknowns when performing the assay. Thus once spacers are added to each primer, depending on the length of the spacers, the empirical length of the resultant product was a different size than expected. Therefore each of the primers once spaced, had to be resized. Table 8b shows the difference between expected and observed primer lengths or peak positions. After the first round of testing, four primers were shown to be in

problematic areas for efficient and robust multiplex interpretation. The tail for primer 2R was increased to 37 nucleotide residues, 55F's spacer was decreased to 10 residues, 62F's spacer was decreased to 31 residues, and 10R's spacer was decreased to 41 residues respectively. These four primers were then re-tested to ensure proper sizing.

Table 9. Finalized SBE mini-multiplex sets 1 through 3 showing color and peak expected peak location.

Gene	Taxa	Primer	Color	Peak Location
Set 1	Ursus	19F	**blue**	23
	Canis	3F	**blue**	54
	Gallus	51F	**green**	29
	Sus	62F	**black**	52
	Human	55F	**black**	30
	Cervidae	13F	**red**	64
Set 2	Gallus	6R	**blue**	37
	Canis	2R	**green**	56
	Felis	23R	**black**	27
	Ursus	13R	**black**	40
	Human	56F	**red**	48
	Mus	10R	**red**	69
Set 3	Felis	14R	**blue**	49
	Sus	26R	**blue**	67
	Cervidae	7R	**green**	42
	Bos	980F	**black**	63
	Bos	841F	**red**	24
	Mus	18F	**red**	36

Once the sizes for each of the spaced primers had been empirically determined the mini-multiplexes were assembled. The primers were placed into the multiplexes in such a way so as to include the fewest primers for each color fluorophore in a single set. This way confounding effects caused by spacing would be less of an issue. Table 9 shows the primers included in each of the multiplexes. Each of the multiplexes were tested on species individually, combinations of two or three species, and on a mixture of all nine species to determine the sensitivity and

accuracy of genotyping. This total species mixture is what in essence serves as the "zoo control" for the assay.

Two Species Mixtures

Several iterations of a two species mixture were set up for identification with the three mini-primer sets. The mixture component species were mixed at both equal and non-equal concentrations, such as may be seen in real world case sample types. Mixtures of 1 to 1, 1 to 10, and 1 to 100 DNA concentrations of white tail deer to domestic cow, human to domestic canine, and black bear to domestic pig were constructed and genotyped. Although there were problems with consistency of amplification and typing across all samples, full profiles for each of the mixtures were recovered for a majority of the samples. Figure 11 shows the results of one of the deer and cow mixtures showing a full profile for both animals. As demonstrated, the mini-set 1 electropherogram exhibits a low background with a strong peak height at the desired location. The electropherogram for the third mini-plex also shows strong peaks at the desired locations; however it also contains both blue and green peaks at the 25 base pair location. These peaks do not appear to be contamination; however they confound the results making the reporting of results from this test presumptive for the presence of the two species rather than confirmatory. This problem is found in most of the samples genotyped with this system. Primer peaks were recovered for samples at 1 to 1 and 1 to 10 concentrations with a proportional loss in RFU peak height. No species were identified at a 1 to 100 concentration.

a.

b.

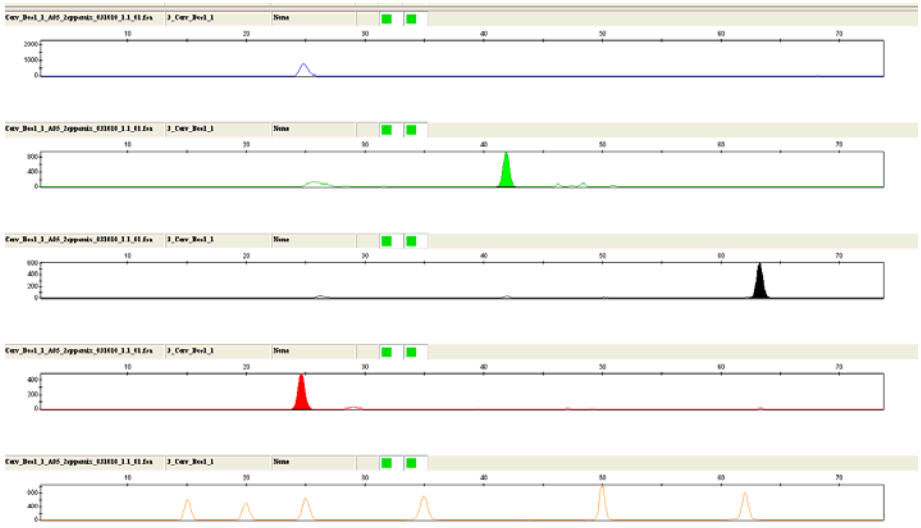


Figure 11. Electropherograms from SBE mini-primer sets 1 (a) and 3 (b) for a two species mixture of deer and cow DNA. Four clear peaks can be seen corresponding to 13F (64 red in set 1); 7R (42 green in set 3); 980F (63 black in set 3); and 841F (24 red in set 3).

Three Species Mixtures

Once it was determined that full profiles could be obtained from two-species mixtures, a three-species mixture of a domestic chicken, domestic feline, and domestic cow in 1:1:1 and 1:10:100 mixture ratios respectively were created and genotyped using the assay. Again there was a lack of consistency with the recovery of alleles; however we did recover full profiles in a number of the trials.

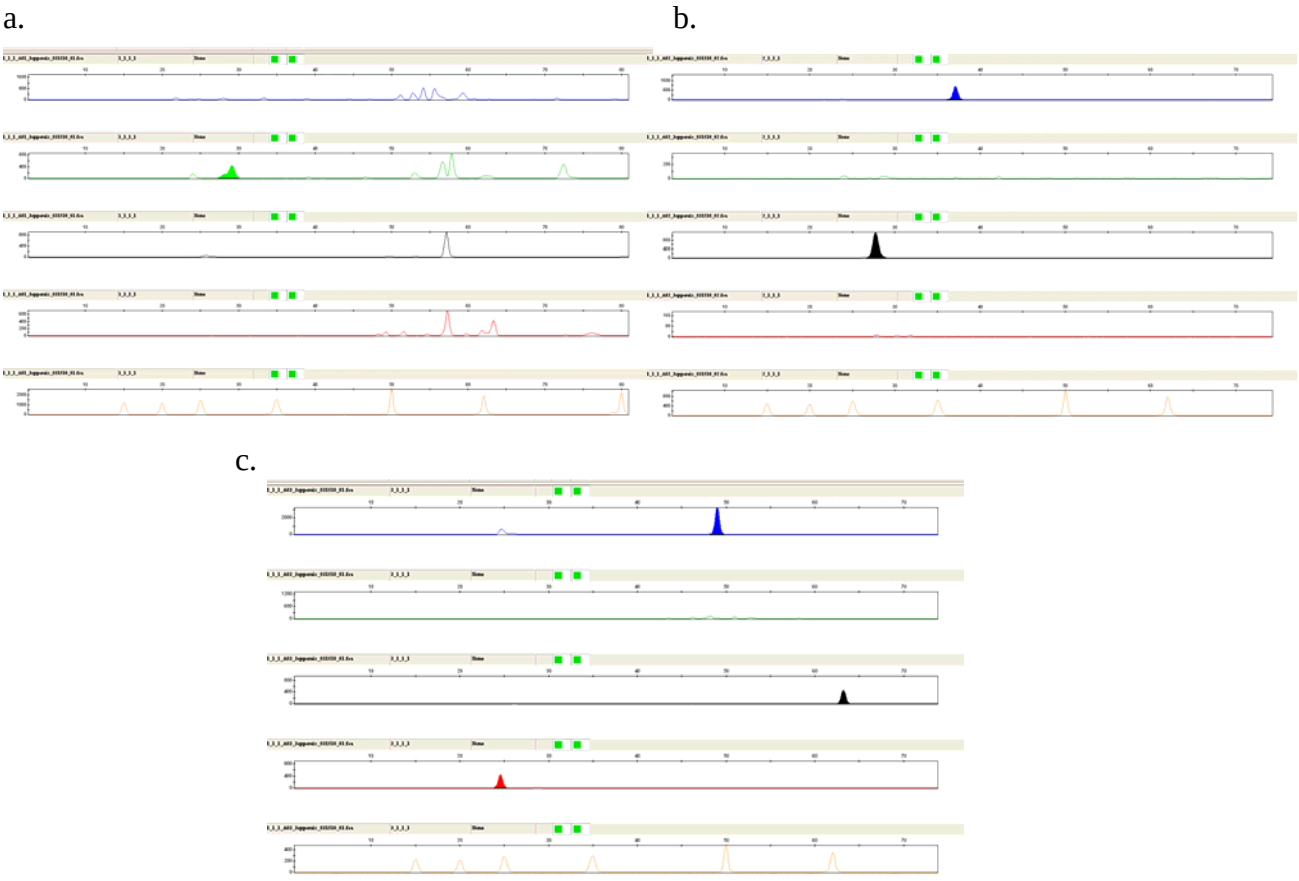


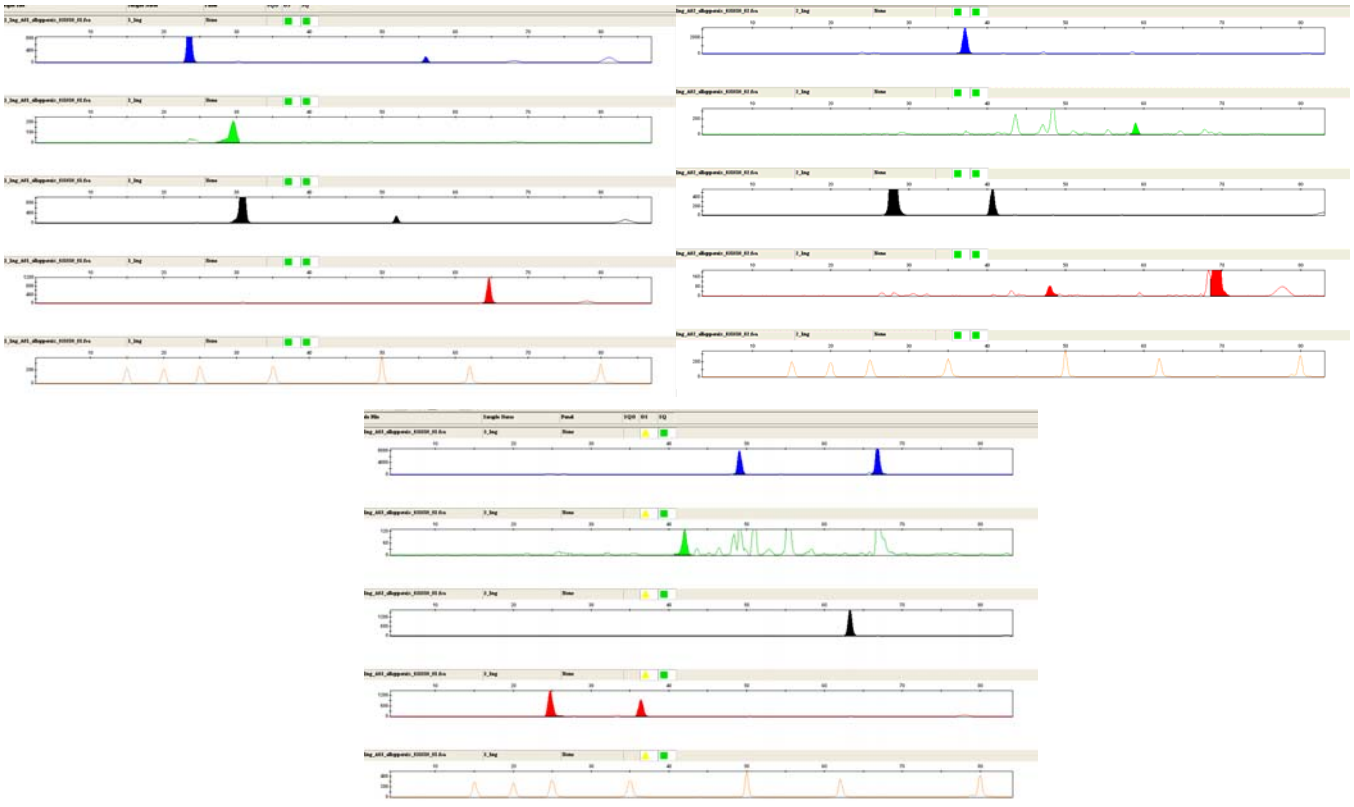
Figure 12. Electropherogram of the three mini-multiplex sets 1 (a), 2 (b), and 3 (c) used to detect a 1:1:1 mixture of domestic chicken, domestic feline, and domestic cow DNA respectively.

At the equimolar mixture (1:1:1) amounts all three individuals were identified, and at the 1:10:100 ratio the only individual not recovered was the domestic cow at a 1:100 rate. Thus we were still able to recover a 1:10 ratio indicating the presence of both the chicken and domestic feline in the mixture. A more sensitive method of detection may detect the cow DNA at a 1:100 level, such a method may be employed after an initial screen with our assay if a the presence of a specific animal is being sought that isn't present in the initial results. In such case, the assay provided was a presumptive test identifying two of the three species in the mixture. Again there were issues with general reproducibility with peaks dropping out altogether for some samples and only partials recovered for others. Also there appeared to be more artifacts in the three species mixture rather than in the two, which can best be noticed in the electropherrogram from primer set 1 in figure 12. There are several artifacts starting after the 50bp mark, though confounding are easy to interpret as artifacts by a technician experienced with analyzing results from this particular assay.

Nine Species Mixture

Lastly in order to determine if it was conceivable to include a "zoo control" which would incorporate all nine of the taxa into one sample that would be carried along with the assay as a positive control, we attempted to get all eighteen peaks to genotype at DNA concentrations of 1ng and .5ng per 2 micro-liters of DNA. In all of these iterations the DNA from each of the animal samples was pre-mixed in equimolar concentrations and aliquots of this mixture were used as templates for the reactions. Results from the total sample mixture tests yielded a wide range of results.

a. b.



c.

Figure 13. Electropherogram of the three mini-multiplex sets 1 (a), 2 (b), and 3 (c) used to detect an equal concentration mixture of 9 individual animals representing the nine taxa identified by the assay.

All in all, a full 18 peak profile was generated from only a handful of the full 9 species samples. The majority of the tests yielded partial profiles with peaks exhibiting a wide range of RFU values. Given that assays performed using the same DNA mixture consisting of the same animal samples in the same volume as a starting template for the reaction and yielded different numbers of recovered alleles, the use of a "zoo control" may be impractical. In the cases where full profiles were generated, there were several problems with the electropherograms. In the sample presented in Figure 13 several loci exhibit extremely low RFU values, two in the first mini-plex (3F - length 54 blue, and 62F - 52 yellow) and one in the second (56F - length 48 red).

These low peaks represent low amplification efficiencies which were not able to be raised through the use of higher SBE primer concentration. Another issue was the presence of spurious alleles and heightened baseline noise. This phenomenon was evidenced in the third mini-plex where primer 7R (length 42 green) butted against a series of spurious peaks, some with RFU heights greater than 400. At the decreased template concentration (.5ng total or approximately 55 picograms) the alleles were proportionally decreased in the samples which did genotype properly. The first and third mini-multiplexes seemed to be the most stable at the lower template amounts, whereas the second mini-plex exhibited extremely low RFU levels for half of the alleles.

Reaction troubleshooting

In order to address the continuing presence of the miscellaneous spurious peaks and background noise, several variables were examined. Applied Biosystems, manufacturer of theSNaPshot™ kit, suggested several possible reasons for the presence of extraneous peaks, including incomplete removal of PCR primers, incomplete removal of ddNTPs, the presence of PCR-amplified templates, and primer dimmers and hairpin formation. Due to the fact that there were a wide range of peak lengths and sizes exhibited by the artifacts confounding the results, each of these aspects could have been a potential causation of our issues. For example in Figure 13c the green lane shows half a dozen artifacts downstream from the expected SNP, exhibiting RFU's too high to attribute to simple background noise. This type of signal to noise ratio is unfortunately commonplace. Hard numbers for the number of noise artifacts per run are tough to develop as the peaks are amorphous and inconsistent and thus difficult to quantify.

All assay primers were created using the Primer 3 program which would have indicated all potential issues with primer self-binding or hairpin formation, thus this was discounted as the

main issue regarding formation of amplification artifacts.. The temperature at which the PCR reaction was performed had the potential to cause unintentional mis-priming, and may have generated a portion of the confounding artifacts. Unfortunately, changes to this temperature greatly affect the ability to generate PCR amplicons across the entire range of species; therefore it cannot be substantially altered.

The troubleshooting tips also suggested that the procedure for reaction clean-up for both the post-PCR and post SNaPshot™ steps, are linked to artifact generation. After PCR, an ExoSap-IT® reaction is used to enzymatically destroy any remaining primer sequence and degrade excess dNTP's simultaneously. Both the addition of more enzymes into the clean-up reaction and lengthening the incubation time were tested. This appeared to decrease a number of the smaller peaks (less than 30bp); however neither change had an effect on artifacts showing up further into the run. The largest artifacts, which are found at or around the 90bp mark, changed little if any. Next, changes to the shrimp alkaline phosphatase reaction (SAP) which is used to clean-up the SNaPshot™ reaction were tested. Again both enzyme levels and length of incubation were altered and little change in spurious peak heights was seen. At present more work is needed to optimize both the PCR and SBE reactions in order to maintain the fidelity of the reactions while reducing the noise.

Conclusions

Through the work performed during the course of this contract, Bode has demonstrated that the use of canonical SNPs to identify the species contributing to either a single source sample or mixture sample is possible. The assay is able to separate two, three, and nine species mixtures using this technology. Technical issues do exist, but once they are resolved the assay could be a useful investigative forensic tool. There are, however, several issues to address before

this assay will be of use to the forensic community at large. In the current form, the assay functions as an initial presumptive screening tool. The scientist performing the assay needs to possess an in-depth understanding of the limitations and issues inherent to the assay in the present form in order to best analyze the results. With a working knowledge of the current assay a scientist should be able to identify the presence of a particular taxon or set of taxa. With the possibility of incurring artifacts, and the possibility of elevated rates of false negative results, there would be no positive identification. This assay can be used now to generate leads and provide probative information, but is not ready to be admitted as evidence. Before an assay like this is ready to be put into widespread use in the forensic DNA analyst's set of tools, more work needs to be performed further optimizing each step of the process. Also it is necessary to examine a much wider number of taxa and individuals than the ninety individuals in nine taxa studied here. Though these SNPs separate the taxa used in this study, there is no way of knowing how the addition of ten, twenty, or one hundred other species will affect the ability of these markers to remain "canonical" to their respective taxa. The development of this first non-cloning based species mixture testing assay provides a strong foundation for future work in this field.

Key Recommendations

As previously mentioned there are a number of areas in which this assay will need to be improved upon. Using the three mitochondrial markers COI, Cytb, and ND4 evidence has been provided that canonical SNPs do exist for use in differentiating the components of a mixture. Although only a small number of usable canonical SNPs was found within these genes, future analyses can be expanded to more mitochondrial genes and eventually the entire mtDNA genome. By identifying more gene candidates, the potential for developing PCR primers

exhibiting more similarity than those used in the creation of this assay will be increased. Exploiting similarity among the primer sets being used in the initial PCR should allow the use of higher, more stringent, annealing temperatures which will substantially decrease the presence of mis-primed amplicons and artifacts.

Also of interest is the expanding of the number of species and taxa in the gene alignments. With the vast amount of mitochondrial sequencing data constantly being uploaded into GenBank, expanding the data matrix could be a reasonably cost effective and rapid method of increasing the number of canonical SNPs identified. Although this particular assay centered on animal taxa commonly encountered in forensic evidentiary samples, the idea may be expanded into different panels. It is possible to create different primer sets for use in particular situations. The more specific the area of interest is, the more likely it is that PCR primers will be found that amplify more efficiently. For instance, a bird panel could be created identifying the major taxa of domestic and endangered birds for use in screening international trade or suspected smuggling cases. Marine panels consisting of marine mammal or marine piscine taxa would also be of great assistance to customs and wildlife officials. This technique can be adopted for any set of animals, and the overall success of the assay should theoretically function better than the total forensic taxa approach taken here.

BIBLIOGRAPHY

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J., Staden, R., and Young, I.G. (1981) Sequence and organization of the human mitochondrial genome. *Nature*, 290 (5806), 457-65.

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., and Howell, N. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet*., 23 (2), 147.

Bataille, M., Crainic, K., Leterreux, M., Durigon, M., and de Mazancourt, P. (1999) Multiplex amplification of mitochondrial DNA fro human and species identification in forensic evaluation. *Forensic. Sci. Int.*, **99**(3), 165-170.

Benson, D.A, Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. (2004) GenBank: update. *Nucleic Acids Res*., **32**(Database issue), *d*23-26.

Bellis, c., Ashton, K.J., Freney L., Blair, B., and Griffiths, L.R. (2003) A molecular genetic approach for forensic animal species identification. *Forensic Sci. Int*., **134** (2-3), 99-108

Blouin, M.S. (1998) Mitochondrial DNA diversity in nematodes. *J. Helminthol*., (72), 285-9.

Branicki, W., Kupiec, T., and Pawloski, R. (2003) Validation of Cytochrome b sequence analysis as a method of speies identification. *J. Forensic Sci*., **48** (1), 83-87.

Butler, J. David, V., O'Brien, S., and Menotti-Raymond, M. (2002) The MeowPlex: A new DNA test using tetranucleotide STR Markers for the domestic cat. Profiles in DNA Vol.5 (2), 7-10.

Carvalho, C.M.B., and Pena, S.D.J. (2005) Optimization of a multiplex minisequencing protocol for population studies and medical genetics. *Genetics and Molecular Research*, **4** (2), 115-125.

Frankham, R., Ballou, J.D., and Briscoe, D.A. (2002) *Introduction to Conservation Genetics*. Cambridge University Press, Cambridge.

Fridez, F., Rochat, S., and Coquoz, R. (1999) Individual identification of cats and dogs using mitochondrial DNA tandem repeats. *Sci. Justice*, **39**(3), 167-71.

Halverson, J.L., and Basten, C. (2005) Forensic DNA identification of animal-derived trace evidence: tools for linking victims and suspects. *Croat. Med. J.* **46**(4), 598-605.

Halverson, J, and Basten, C. (2005) A PCR multiplex and database for forensic DNA identification in dogs. *J. Forensic. Sci.*, **50** (2), 352-363.

Herbert, P.D.N., Stoeckle, M.Y., Zemlak, T.S., and Francis, C.M. (2004) Identification of birds through DNA barcodes. *PLoS Biology*, **2** (10), 1657-1663.

Inagaki, S., Yamamoto, Y., Doi, Y., Takata, I., T., Imabayashi, K., Yoshitome, K., Miyaishi S., and Ishizu, H. (2004) A new 39-plex analysis method for SNPs including 15 blood group loci. *Forensic Sci. Int*., **144** (1), 45-57.

Kaderali, L., Deshpande, A., Nolan, J.P., and White, P.S. (2003) Primer-design for multiplexed genotyping. Nucleic Acids Research, 31 (6), 1796-1802.

Liu, J., Berry, R.E., and Blouin, M.S. (1999) Molecular differentiation and phylogeny of entomopathogenic nematodes (rhabditida; heterorhabditidae) based on ND4 sequences of mitochondrial DNA. *J. Parasitol*., 85 (4), 709-15.

Menotti-Raymond M.A., David, M.A., and O'Brien S.J. (1997) Pet cat hair implicates murder suspect. *Nature*, **386** (6627), 774.

Onofri, V., Alessandrini, F., Turchi, C., Pesaresi, M., Buscemi, L., Tagliabracci, A. (2006) Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs. *Forensic Sci. Int.* **157** (1), 23-35.

Pfeiffer I, Burger J, Brenig B (2004) Diagnostic polymorphisms in the mitochondrial cytochrome b gene allow discrimination between cattle, sheep, goat, roe buck and deer by PCR-RFLP. *BMC Genetics*, **5**, 30

Sanchez, J. J., Borsting, C., Hallenberg, C., Buchard, A., Hernandez, A., and Morling, N. (2003) Multiplex PCR and mini-sequencing of SNPS-a model with 35 Y chromosome SNPs. *Forensic Sci. Int.*, **137** (1), 74-84.

Sanchez, J. J., Borsting, C., and Morling, N. (2005) Typing of Y chromosome SNPs with multiplex PCR methods. *Methods Molecular Biology,* **297,** 209-228.

Savolainen, P., Arvestad, L., and Lundeberg, J. (2000) A novel method for forensic DNA investigations: repeat-type sequence analysis of tandemly repeated mtDNA in domestic dogs. *J. Forensic Sci*,. **45**(5), 990-999.

Savolainen, P., and Lundeberg J (1999) Forensic evidence based on mtDNA from dog and wolf hairs. *J. Forensic Sci.*, **44**(1), 77-81.

Shutler, G.G., Gagnon, P., Verret, G., Kalyn, H., Korkosh, S., Johnston, E., et al. (1999) Removal of PCR inhibitor and resolution of DNA STR types in mixed human-canine stains from a five year old case. *J. Forensic Sci*., **44**(3),623-626.

Yoon, C.K. (1993) Botanical witness for the prosecution. *Science*, **260** (5110), 894-895.

Appendix A

COI Forward Strand Primers

Taxa	Position	Change	Color	SBE primer sequence 5'-3'
Canis				
1	396	A > G	blue	GAACCGTATACCCCCCACT
2	586	C > T	red	AATTACAGCAGTTCTACTC
3	606	A > G	blue	ACTATCCCTGCCTGTACT
Felis				
4	334	C > T	red	CCCTCCATCCTTTCTACTC
Hominid				
5	313	C > T	red	AACATAAGCTTCTGACTC
6	325	T > C	black	CTGACTCTTACCTCCCTCT
7	407	T > A	green	CCCTCCCTTAGCAGGGAACT
8	409	G > T	red	CCCTTAGCAGGGAACTAC
9	415	G > C	black	GCAGGGAACTACTCCCAC
Gallus				
10	350	T > C	black	CTCCTACTAGCCTCATCTA
11	526	A > C	black	AACATAAAACCCCCCGCA
12	577	G > A	green	ATCCGTCCTCATTACTGCC
Cervidae				
13	643	C > T	red	CTATTAACAGACCGAAAT

COI Reverse Strand Primers

Taxa	Position	Change	Color	SBE primer sequence 5'-3'
Canis				
1	396	T>C	black	CATGGGCCAGATTGCCAGC
2	586	G>A	green	TAATACAGGCAGGGATAGTA
3	606	T>C	black	GTATTGTAATTCCAGCAGC
Felis				
4	334	G>A	green	TACCATAGATGAGGCGAGTA
Hominid				
5	313	G>A	green	GAGTAGGAGAGAGGGAGGTA
	325	A>G	blue	ACATGCGAGCAGGAGTAGGA
	407	A>T	red	CGGAGGCTCCAGGGTGGGAG
	409	C>A	green	TACGGAGGCTCCAGGGTGGG
	415	C>G	blue	TAGGTCTACGGAGGCTCCAG
Gallus				
6	350	A>G	blue	TGCCGGCCCCAGCTTCTACG
	526	T>G	blue	GGGTGTTTGGTATTGTGACA
	577	C>T	red	TAAGGAGAGGAGTAGTAGGA
Cervidae				
7	643	G>A	green	CGAAAAAGGTTGTGTTTA

CytB Forward Strand Primers

Taxa	Position	Change	Color	SBE primer sequence
Gallus				
29	265	A > T	red	CACGCAAACGGCGCCTCA
30	344	T > C	black	ACAAGGAAACCTGAAACA
31	405	A > G	blue	GGGCTATGTTCTCCCATG
32	472	A > C	black	GCAATTCCCTACATTGGA
33	550	A > C	black	TTCTTCGCTTTACACTTC
34	572	C > G	blue	TCCCCTTTGCAATCGCAG
Mustelid				
35	397	C > T	red	GCATTCATAGGTTACGTC
36	355	C > T	red	TGAAACATCGGCATCATC
Sus				
37	366	A > C	black	AGTAGTCCTACTATTTAC
Hominid				
38	320	A > T	red	GCCTATATTACGGATCAT
39	445	C > T	red	GCCACAGTAATTACAAAC
40	484	G > C	black	ATTGGGACAGACCTAGTT
41	503	T > A	green	AATGAATCTGAGGAGGCT
42	515	A > G	blue	GAGGCTACTCAGTAGACA
43	538	G > A	green	ACCCTCACACGATTCTTT
Bos				
63	841	C>T	red	TTATTTGCATACGCAATC
64	980	T>C	black	GCCAATGCCTATTCTGAG

CytB Reverse Strand Primers

Taxa	Position	Change	Color	SBE primer sequence
Gallus				
16	265	T>A	green	GAAGATACAGATGAAGAAGA
	344	A>G	blue	TGAGGAGGAGGATTACTCCT
	405	T>C	black	CCTCAGAATGATATTTGGCC
	472	T>G	blue	GGCTCACTCTACTAGGGTGT
	550	T>G	blue	TGCGATTGCAAAGGGGAGGA
	572	G>C	black	TGAGGTGGATGATAGTAATA
Mustelid				
17	397	G>A	green	GATATTTGTCCTCATGGTA
18	355	G>A	green	TGCTATTACTGTGAATAGTA
				TGCTATGACTGTGAATAACA
Sus				

19	366	T>G	blue	ATGAAGGCTGTTGCTATAAC
Hominid				
20	320	T>A	green	TTCAGGTTTCTGAGTAGAGA
	445	G>A	green	GTATGGGATGGCGGATAGTA
	484	C>G	blue	GTAGCCTCCTCAGATTCATT
	503	A>T	red	GGGTGGGACTGTCTACTGAG
	515	T>C	black	AGAATCGTGTGAGGGTGGGA
	538	C>T	red	GGGCAAGATGAAGTGAAAGG
Bos				
27	841	G>A	green	TGTTGGGGATTGATCGTA
28	980	A>G	blue	AGTAGGTCTGCTACTAGG

Cytb Universal Forward Strand Primers

Taxa	Position	Change	Color	SBE primer sequence
Gallus				
44	113	G > C	black	ATTTCGGCTCCCTATTAG
45	128	T > C	black	AGCAGTCTGCCTACATGA
46	148	T > C	black	CAAATCCTCACCGGCCTA
47	181	A > C	black	TACACAGCAGACACATCC
48	182	C > T	red	ACACAGCAGACACATCCC
49	206	T > C	black	TCTCCTCCGTAGCCCACA
50	214	G > A	green	GTAGCCCACACTTGCCGG
51	241	T > A	green	TACGGCTGACTCATCCCGG
52	265	A > T	red	CACGCAAACGGCGCCTCA
53	344	T > C	black	ACAAGGAAACCTGAAACA
Hominid				
54	116	T > C	black	TCGGCTCACTCCTTGGCG
55	137	T > C	black	GCCTGATCCTCCAAATCA
56	166	A > T	red	TTCCTAGCCATGCACTAC
57	175	A > G	blue	ATGCACTACTCACCAGAC
58	177	A > C	black	GCACTACTCACCAGACGC
Felis				
59	153	A > G	blue	CCTCACCGGCCTCTTTTT
Mustelid				
60	355	C > T	red	TGAAACATTGGTATTGTC
61	397	C > T	red	AGCATTCATAGGTTACGTA
Sus				
62	366	A > C	black	AGTAGTCCTACTATTTAC

Cytb Universal Forward Strand Primer

Taxa	Position	Change	Color	SBE primer sequence
Gallus				
21	113	C>G	blue	TTTGGGTCATGAGGCAGACT
	128	A>G	blue	GTAGGCCGGTGAGGATTTGG
	148	A>G	blue	TGTGTAGTGCATGGCTAGTA
	181	T>G	blue	GGCTACGGAGGAGAAGGCTA
	182	G>A	green	GGGCTACGGAGGAGAAGGCT
	206	A>G	blue	CGTATTGTACGTTCCGGCAA
	214	C>T	red	GAGTCAGCCGTATTGTACGT
	241	A>T	red	GGCGCCGTTTGCGTGGAGAT
	265	T>A	green	GAAGATACAGATGAAGAAGA
	344	A>G	blue	TGAGGAGGAGGATTACTCCT
Hominid				
22	116	A>G	blue	TGATTTGGAGGATCAGGCAG
	137	A>G	blue	TGGCTAGGAATAGTCCTGTG
	166	T>A	green	GGCGGTTGAGGCGTCTGGTG
	175	T>C	black	TGATGAAAAGGCGGTTGAGG
	177	T>G	blue	ATTGATGAAAAGGCGGTTGA
Felis				
23	153	T>C	black	TCTGATGTGTAGTGTATGGC
Mustelid				
24	355	G>A	green	TGCTATTACTGTGAATAGTA
25	397	G>A	green	GATATTTGTCCTCATGGTA
Sus				
26	366	T>G	blue	ATGAAGGCTGTTGCTATAAC

ND4 Forward Strand Primers

Taxa	Position	Change	Color	SBE primer sequence
Hominid				
14	706	C > T	red	GTACTTGCCGCAGTACTC
15	938	A > G	blue	TGAAGCTTCACCGGCGCA
Mus				
16	634	C > G	blue	AAAATACCATTATATGGA
17	718	G > A	green	ATTCTTCTAAAACTAGGT
18	739	A > T	red	TACGGAATAATTCGCATC
Ursus				
19	752	A > G	blue	GAATTACAACACTACTTG
20	764	A > G	blue	TACTTGGTCCTTTAACGA
Felis				
21	759	T > C	black	CAGTCCTACTTAACCCCA

Gallus				
22	761	A > T	red	CTCCTAATGGAGCCCGTA
23	769	A > T	red	GGAGCCCGTATCCAACTT
24	772	G > C	black	CCCGTATCCAACTTCTTA
25	788	T > C	black	TGGCCTACCCCTTCATAA
26	805	A > G	blue	ACCCTAGCCCTATGAGGT
27	884	C > G	blue	CATCCGTAAGCCATATGG
28	902	T > G	blue	GCCTAGTCATCGCTGCAA

ND4 Reverse Strand Primers

Taxa	Position	Change	Color	SBE primer sequence
Hominid				
8	706	G>A	green	ACCATAGCCGCCTAGTTTTA
	938	T>C	black	CCGTGGGCGATTATGAGAA
Mus				
9	634	G>C	black	TTTTGGTAGTCATAGGTGAA
10	718	C>T	red	GATGCGAATTATTCCGTAAC
11	739	T>A	green	TAGTGGGTCTAGAATAATGG
Ursus				
12	752	T>C	black	CCATGAAGCTTGTTAGAGGG
13	764	T>C	black	TGAAGGGGTAAGCCATGAAG
Felis				
14	759	A>G	blue	GGTATGCCATTTGGTTCGTT
Gallus				
15	761	T>A	green	GGGGTAGTGTAAGAAGTTGG
	769	T>A	green	GGTGAGGAAGGGGTAGTGTA
	772	C>G	blue	TAGGGTGAGGAAGGGGTAGT
	788	A>G	blue	GGGCACCTCATAGGGCTAGG
	805	T>C	black	GATAGAGCTAGTTATTAGGG
	884	G>C	black	TACTTGCAGCGATGACTAGG
	902	A>C	black	ATCATTGGGTTTGGATTATA