This is to certify that the
thesis entitled

VALIDATION OF CYTOCHROME B PRIMERS FOR
FORENSIC SPECIES IDENTIFICATION

presented by

Sherri Lindamarie Freeman

has been accepted towards fulfillment
of the requirements for the

M.S. degree in Criminal Justice

Major Professor's Signature

April 20, 2004

Date

VALIDATION OF CYTOCHROME B PRIMERS FOR FORENSIC SPECIES IDENTIFICATION

By

Sherri Lindamarie Freeman

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Criminal Justice

2004

ABSTRACT

VALIDATION OF CYTOCHROME B PRIMERS FOR FORENSIC SPECIES DIFFERENTIATION

By

Sherri Lindamarie Freeman

The mitochondrial DNA section of the Armed Forces DNA Identification Laboratory (AFDIL) is primarily responsible for the analysis and characterization of ancient remains received from the Central Identification Laboratory in Hawaii. The specimens received by the mitochondrial DNA analysts have been exposed to varied environmental conditions and can be between 40–60 years old. At times, specimens are so small or degraded that they cannot be anthropologically distinguished as human or non-human. This becomes an issue when the degraded nature of the DNA and human specificity of the control region primers used by the scientists prevents determination of the cause(s) of amplification failure.

This thesis is a validation study that was undertaken to provide a procedure for species identification by amplification, sequencing, and either BLAST or phylogenetic comparison to identify species. The mitochondrial cytochrome b gene was chosen because of its known success for species differentiation and the existence of optimized universal primer sequences. Validation of the technique involved amplification optimization, sensitivity and specificity studies, comparison of identification methods, and mixture analysis.

ACKNOWLEDGMENTS

I would like to express my appreciation to the scientists at the Armed Forces DNA Identification Laboratory for allowing me the wonderful educational opportunity that my internship and graduate research position provided. I would especially like to thank Timothy McMahon, Ph.D. for all of his guidance and for all of the knowledge that I was able to gain from working with him. I would also like to acknowledge my professors at Michigan State University, Drs. David Foran and Jay Siegel. It would take a lot more than words for Dr. Siegel to realize how much the Michigan State program helped me accomplish, and I am ever grateful for the opportunity to be a part of it. To Dr. Foran, I appreciate your patience also and your persistence in helping me to complete this project. I know it was a challenge collaborating on this from a distance, but I have learned a lot from the experience. Last but not least, I would like to express my deepest gratitude to all of my friends, my family, and my fiance who stood behind me and cheered me on through the whole thing.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The use of DNA in forensic analysis has advanced steadily over the past two decades with the introduction of newer, faster, more reliable methods to aid in criminal investigations. Current DNA methodologies are constantly re-evaluated to find ways to enhance such aspects as their associated instrumentation, their robustness, and their accuracy. Continued improvements to DNA methodologies are necessary to aid in correctly identifying perpetrators in rape cases, fathers in paternity cases, remains from homicide and missing persons cases, and trace biological material associated with various crime scenes.

One area that has been continually targeted for improvement is species differentiation, which is most often used in wildlife forensic cases. In this context, the field has evolved from the use of protein-based differentiation methods to polymerase chain reaction (PCR) based short tandem repeat (STR) and mitochondrial DNA (mtDNA) assays for individual species identification. Given the need for species determination of trace biological remains in human criminal cases, over the past decade forensic scientists have begun to draw upon, enhance, and validate wildlife forensic species differentiation techniques. For example, in cases where a criminal enters a person's home, hairs from the victim's pet cat or dog may be available to link a suspect to the crime. An example is the MeowPlex, a system that allows identification of the source of cat hair using felid-specific nuclear STR markers (Butler et al. 2002). The MeowPlex uses 11 STR markers chosen based on analysis in 37 different breeds of cat common to the United States. It is different from other species identification techniques currently in use because it is not only species-specific but also helps match the hairs to a particular cat (depending upon

the breed) in much the same way human identity testing using STR markers enables unique identification of human DNA specimens (Butler et al. 2002). Although the MeowPlex may only work on certain breeds and has yet to attain the identity statistics achievable with human STR kits, it has laid the foundation for the development of improved cat STR testing kits and kits for other common domestic animals (Butler et al. 2002).

The MeowPlex is one of the most recent developments in an ongoing effort to improve the efficiency and effectiveness of DNA-based species identification methods that commenced in the late 80s and early 90s, after PCR was introduced into forensics laboratories. Advancements associated with these efforts include improvements in testing kits, reagents, equipment, and the instrumentation associated with DNA extraction, amplification, sequencing, and analysis. These methods have helped bring mtDNA analysis to the forefront in the quest for improved species differentiation procedures.

The improvements in extraction and amplification methods coupled with the apparent resiliency of mtDNA allow for mtDNA to be isolated from highly degraded remains even when nuclear DNA testing techniques have failed (Holland et al. 1993). Scientists have yet to determine the exact reason why mtDNA can be obtained from highly degraded material, but three theories have been presented. The first is based on the fact that cells typically have between 900 and 1300 mtDNA molecules compared to the single copy nuclear genome (Bogenhagen and Clayton 1974, Moraes et al. 1999, reviewed by Scheffler 1999, Veltri et al. 1990). The second is that the double-stranded, closed, circular nature of the mtDNA allows it to withstand environmental and cellular

agents that degrade nuclear DNA. The third is that mtDNA is protected within the mitochondrion, whose membrane may be much more resilient than the nuclear membrane.

The Mitochondrion: Structure and Function

Mitochondria are thought to have arisen from small, rod-shaped eubacteria that survived in a symbiotic relationship with anaerobic, unicellular eukaryotes that engulfed the eubacteria and utilized their aerobic respiratory capabilities. Scientists speculate that the eubacteria were eventually incorporated into the cell where they retained their respiratory capabilities but lost their ability to function independently (reviewed by Scheffler 1999). These eubacteria became the ancestral version of the present day mitochondrion. As the mitochondria evolved, a portion of their DNA was retained and is now the mitochondrial genome while the remainder of the eubacterial DNA was either eliminated or exported to the nucleus, (reviewed by Shadel and Clayton 1997). Eventually, the respiratory capabilities provided by the mitochondrial and nuclear-encoded proteins gave rise to the oxidative phosphorylation pathway, which provides cells with the energy needed to survive. Many key proteins of this pathway are encoded by the portion of the eubacterial DNA that evolved into the mtDNA genome, with several of the essential accessory proteins supplied by the eubacterial genes that were incorporated into the nuclear genome. The cellular respiration pathways are well conserved among most organisms and produce adenosine triphosphate (ATP), which acts as an energy carrier or transporter that: (i) drives the functional processes of the mitochondrion as well as other cellular organelles, (ii) provides enough energy to drive specific bodily functions (e.g., muscle contraction and sperm motility), and (iii) maintains

the body temperature of warm-blooded organisms (reviewed by: Alberts 2002, Lewin 1998, Scheffler 1999). For a more comprehensive review of the functions of the mitochondrial genome, refer to Lewin (1998) and Scheffler (1999).

Although the respiratory functions of the mitochondrion are highly conserved between vertebrates and invertebrates, their genome sizes and gene orders are not, even though they encode many of the same basic structures (e.g., tRNAs, rRNAs, cytochrome oxidases, etc.; Roe et al. 1985, reviewed by Scheffler 1999). The mitochondrial genomes of invertebrates are structured much like mammalian nuclear genomes, having numerous introns (some transposable) and noncoding regions, making their genomes larger and more complex than vertebrate mtDNA genomes (Nobrega and Tzagoloff 1980, review by Scheffler 1999).

Vertebrate mitochondrial genomes, on the other hand, are smaller, and the majority of the DNA codes for proteins. For example, the human mitochondrial genome is 16,569 nucleotides in length and all but the 1122 nucleotides of the control region is coding (Figure 1) (Anderson et al. 1981, reviewed by Alberts et al. 2002, Scheffler 1999). Although the mtDNA control region does not encode any proteins, it does contain two transcriptional promoters, the light strand promoter (LSP) and the heavy strand promoter (HSP), as well as the heavy strand origin of replication (Anderson et al. 1981, reviewed by Scheffler 1999, Shadel and Clayton 1997). The light strand origin of replication on the other hand is located near the Cox I gene (Figure 1). Replication commencing from the heavy strand origin of replication is especially notable because of the possible formation of a D-loop, which arises from a newly synthesized heavy strand segment and the original heavy strand template (Arnberg et al. 1971, Chang and Clayton 1985,

reviewed by Alberts 2002, Scheffler 1999, Shadel and Clayton 1997). The control region is of great interest for evolutionary studies because it has a high rate of mutation (Stoneking et al. 1991). The region is also useful for species differentiation because the high rate of intraspecies variation can be combined with the lower rate of mutation of the adjacent cyt b region (discussed below) and adjacent tRNA genes.

The control region also contains two hypervariable regions (HVI and HVII) and two variable regions (VRI and VRII) (Figure 1). Although certain segments in the HV regions are highly conserved (e.g. - conserved sequence blocks and the RNAse MRP cleaving site), overall both HV regions and VRs exhibit a higher degree of sequence substitutions among species and non-maternally related individuals than is observed within the mitochondrial genome as a whole (Grzybowski 2000, Meyer et al. 1999, Parsons et al. 1997). Taking into account both the mutation rate within the control region and the mutation rate of the rest of the genome, the mitochondrial genome mutates at approximately 3.4×10^{-7} bases per generation, or about ten times the rate for the coding regions of nuclear DNA (Brown et al. 1979, Jobling et al. 2004). MtDNA's higher mutation rate in conjunction with its maternal (unilateral) inheritance makes it a prime candidate for delving into evolutionary events. The high mutation rate is beneficial because even the most conserved mitochondrial genes have sufficient sequence differences to allow evolutionary changes to be easily identified (Honeycutt et al. 1995, Ingman et al. 2000, Irwin et al. 1991, Johns and Avise 1998). The unilateral inheritance of mtDNA is advantageous because heterozygosity is not a factor when conducting sequencing studies, making mutations easier to follow from generation

Figure 1. The Mammalian Mitochondrial Genome. The variable regions are indicated in dark gray (hypervariable regions) and light gray (variable regions). Note the positions of the transcription promoters (P_H, P_L) and the heavy strand origin of replication (O_H) within the control region and the position of the cytochrome b gene immediately adjacent to the threonine tRNA gene (THR) to the right of the control region. Also, note the position of the light strand origin of replication (O_L) on the opposite side of the genome. Figure from Lehtonen 2002.

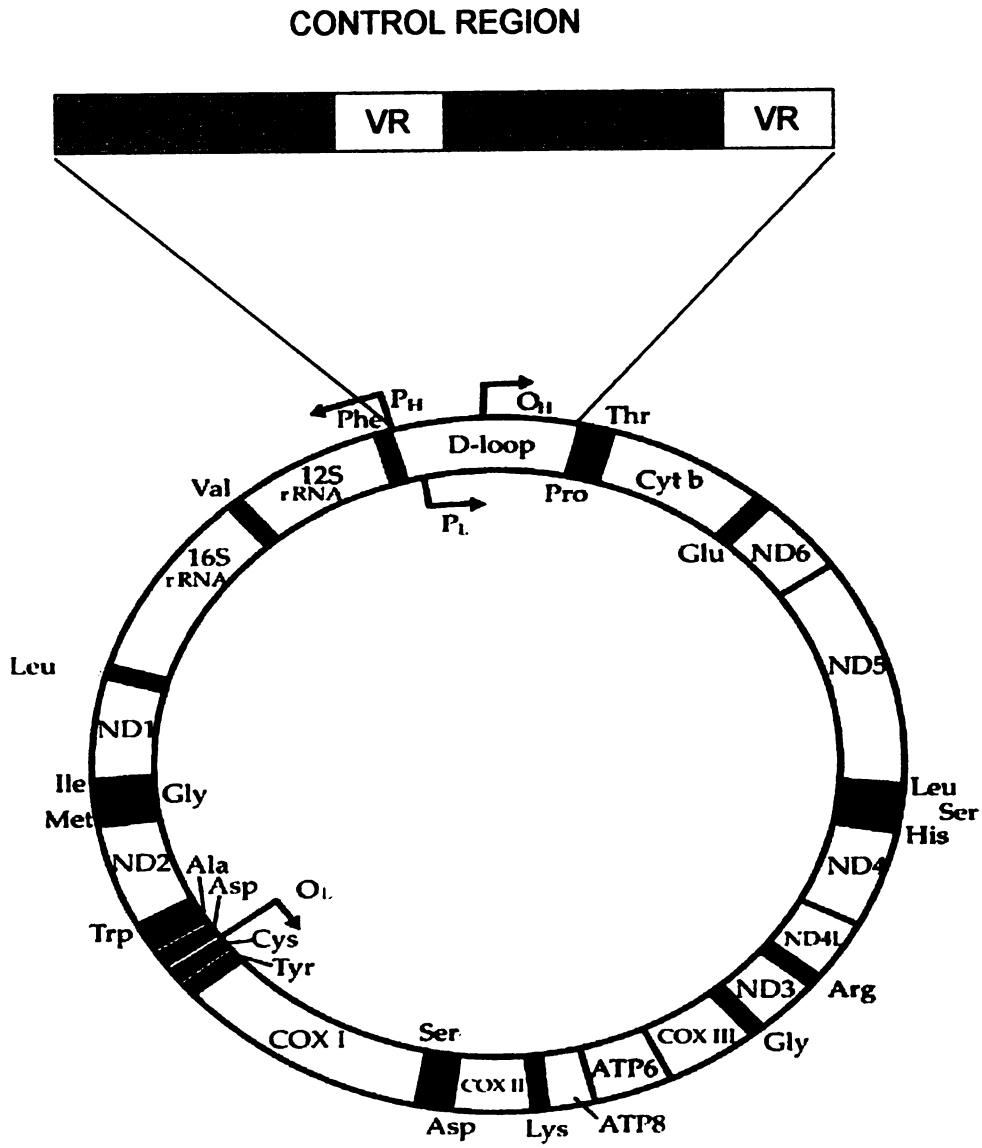

CONTROL REGION
VR
VR
Phe
PH
OH
Thr
D-loop
Pro
Cyt b
Glu
ND6
ND5
Leu
Ser
His
ND4
ND4L
Arg
ND3
Gly
COX III
ATP6
ATP8
Lys
Asp
COX II
Ser
COX I
Trp
Cys
Tyr
Ala
Asp
OL
ND2
Met
Ile
Gly
ND1
Leu
16S rRNA
Val
12S rRNA
PL

Fig. 1. (cont.)

to generation (Giles et al. 1980, Hutchison et al. 1974).

MtDNA and Species Differentiation in Forensics

Current mtDNA testing techniques for human identification use human specific primers to amplify and sequence the HVI and II regions and the VR regions when needed (Sullivan et al. 1992, Wilson et al. 1993 and 1995). The derived sequences are then compared to the Cambridge reference sequence (modified from Anderson et al. 1981) to identify any variations from the Cambridge reference. The identified variations determine an individual's haplotype, which can then be compared to a direct reference from the individual or from a sibling or other maternal relative to see if they match (Holland et al. 1993, Wilson et al. 1993 and 1995, reviewed by Holland and Parsons 1999).

At times, the human specific primers fail to amplify the extracted DNA, which may occur because the extract contains PCR inhibitors, highly degraded DNA, low copy number, or because a non-human template was used (Holland et al. 1993). To overcome PCR inhibition, the DNA template is usually diluted and amplified with an increased volume of Taq DNA polymerase or re-cleaned through further organic extraction or by using purification columns. In instances where the DNA is highly degraded, an increased volume of extract may be amplified with the original primers using more PCR cycles, or the extract may be amplified with primers targeting a shorter sequence segment. Increases in cycle number and/or volume of extract are also utilized when low copy number is encountered. However, if amplification was unsuccessful because a non-human template was used, most DNA forensic laboratories waste valuable time and resources attempting to pinpoint the problem because they do not have a validated

method for identifying the species. Therefore, the development of an efficient species identification procedure could save time and resources for forensic DNA laboratories by helping to elucidate the reason(s) for unsuccessful amplification of remains.

Original wildlife forensic species differentiation based on molecular methods included protein-based assays such as western blotting, enzyme-linked immunosorbent assays (ELISAs), and high performance liquid chromatography (HPLC) analysis. Although effective, these methods had inherent flaws because they often required larger sample volumes than could be obtained from degraded remains and were sensitive to protein degradation issues (Espinoza et al. 1996, Kang et al. 2003, Sarkioja et al. 1988). Another disadvantage was that antibodies from closely related species could cross-react, making accurate species interpretation difficult (Iwasa 1982). To address degradation and cross-reactivity issues, DNA based tests, such as restriction fragment length polymorphism (RFLP) analysis, were developed for species identification. RFLP, one of the earliest DNA techniques used in forensic identification (Cronin et al. 1991, Blackett and Keim 1992, Guglich et al. 1994), utilized one or more restriction enzymes to cut DNA at certain sites within a sequence. Though effective, the procedure requires a good deal of time and large amounts of blood or tissue to obtain sufficient amounts of DNA for testing (Blackett and Keim 1992, Cronin et al. 1991, Guglich et al. 1994). Other challenges include the generation of identical banding patterns with different species or generation of different banding patterns because of heterozygosity in one individual (Guglich et al. 1994).

When identical banding patterns or heterozygosity are encountered, they can only be addressed by performing RFLP analysis with additional restriction enzymes or by

analyzing a different DNA segment (Blackett and Keim 1992, Guglich et al. 1994). The need for additional restriction data also becomes a problem because of the large amount of time required for development of database reference samples for each enzyme (Blackett and Keim 1992, Cronin et al. 1991, Foran et al. 1997b, Guglich et al. 1994, Meyer et al. 1995). For example, both Cronin et al. (1991) and Blackett and Keim (1992) had to use additional restriction enzymes to distinguish among deer species when identical banding patterns were obtained after the initial digestion. Even after restriction digestion of mtDNA with several enzymes, indistinguishable banding patterns were present for some closely related species (Cronin et al. 1991). When this occurred, immunological assays or assessment of other genetic markers was required for differentiation of deer species, including analyses of a serum albumin marker. Once again, though effective, the time needed to conduct additional studies was a factor.

Many of the constraints seen with protein-based species determination were eliminated with the introduction of PCR into DNA forensics, including some RFLP-based methods. Use of PCR for species identification can be applied to either mtDNA or nuclear DNA (Foran et al.1997a and b, Kocher et al. 1989, Naito et al. 1992, Ono et al. 2001, Parson et al. 2000, and Rajapaksha et al. 2002). Though there are multiple DNA regions in both the nuclear and mitochondrial genomes that can be used for species identification, only two of the most commonly used mtDNA segments, the cyt b gene and the control region, will be discussed here. In some instances the control region has been used in conjunction with cyt b for species identification (Bellis et al. 2003, Foran et al. 1997a and b). Foran et al. (1997a and b) used universal vertebrate primers to identify the species of DNA extracted from hair, scat, and other tissues (blood, ear clip, etc.) from 14

North American carnivore species. These primers amplify an approximately 600 bp region of the control region and the 5' end of the cyt b gene from as little as 0.01 µl of extracted DNA using 35 amplification cycles. Agarose gel band sizes were used for initial species differentiation, and RFLP analysis was used for identification of those species that could not be distinguished by agarose gel banding sizes alone. However, in addition to previously mentioned drawbacks, the size (~600 bp) of the target region may prevent complete amplification in cases of DNA degradation.

For degraded specimens where vertebrate specific primers targeting a combined cyt b and D-loop region were ineffective, vertebrate primers that amplified a 300 to 500 bp region of cyt b were developed (Kocher et al. 1989, Rajapaksha et al. 2002, Wetton et al. 2002). The role of cyt b as one of the essential proteins involved in Complex III of the electron transport chain was beneficial because it results in sequence length conservation among vertebrates (Bose et al. 2003, Kocher et al. 1989, reviewed by Scheffler 1999). In contrast, other vertebrate mitochondrial genes, such as ATPase 6, vary in length (Bose et al. 2003, reviewed by Scheffler 1999). In addition, the cyt b gene, like the control region, inherently possesses the beneficial qualities of mtDNA including high copy number, high mutation rate, maternal inheritance, etc. Furthermore, the numerous applications and studies documented through the literature provide an excellent practical foundation for why the cyt b gene has been targeted as a successful candidate for species differentiation (Bartlett and Davidson 1992, Irwin et al. 1991, Kocher et al. 1989, Parson et al. 2000, Scheffler 1999).

One of the earliest sets of universal cyt b primers was developed by Kocher et al. (1989). These authors analyzed published cyt b gene sequences of cow, human, fly and

frog to identify conserved regions. From these, a set of universal primers (L14841, H15149, based on the numbering of the human mitochondrial genome) was developed that would amplify approximately 348 bp (including primer sequences) of the 5' end of the vertebrate cyt b gene. One potential drawback of universal primers is that there may be amplification efficiency problems when analyzing vertebrate samples that have sequence differences within the primer binding site. However, since 1989, modified versions of Kocher et al.'s (1989) primers have been used in a number of species identification studies, including those of Branicki et al. (2003), Hsieh et al. (2003), and Parson et al. (2000). Parson et al.(2000) performed a validation using primers with 9 bases removed from the 5' ends of Kocher et al.'s (1989) original forward and reverse primers. With these modified primers, the authors were able to amplify DNA from the 44 vertebrate species tested, including problematic specimens such as hair bristles and bone extracts, using 30 or 35 PCR cycles. The amplified specimens were identified by phylogenetic comparison, which involves the comparison of specific characteristics, or character states, to determine evolutionary relationships among organisms based on similarities or differences. For DNA comparison, the character state is the DNA sequence for a specific segment under study.

Parson et al. (2000) used the basic alignment search tool (BLAST), discussed in detail below, for their phylogenetic analyses. The same set of primers were used in a BLAST based study by Branicki et al. (2003). Using 32 or 36 amplification cycles (depending on the tissue), the group was able to achieve a sensitivity of 5 pg total DNA and could identify all but three of thirty-four vertebrate species with BLAST.

Hsieh et al. (2001 and 2003) used Kocher et al.'s (1989) reverse primer with Irwin

et al.'s (1991) forward primer to amplify a 402 bp segment of the cytochrome b gene, which was used and for phylogenetic comparison of several species of rhinoceros with Holstein cow and to identify unknown samples. This set of primers was used after amplification of the full ~1100 bp cyt b gene failed to produce a product. Likewise, species specific cyt b primers have been used to detect the presence of protected or endangered animal matter in processed or powdered samples when investigating poaching and illegal trade practices (Meyer et al. 1995; Wan and Fang 2003, Wetton et al. 2002). Wan and Fang (2003) developed a set of tiger specific cyt b primers for regulation of the sale of tiger meat. These primers were successfully used to amplify and identify a single hair as well as dried skin and a specimen of decayed meat. Wetton et al. (2002) developed a different set of tiger specific cyt b primers to determine whether the animal matter in traditional Chinese medicines was from an endangered tiger species. The specimens presented a challenge because the animal bone had been boiled and powdered. The successes of these and other studies demonstrate that cyt b primers are effective for low copy number and/or degraded DNA specimens and that phylogenetic analysis is an effective tool for species determination using the cyt b gene.

Analyses used for identification of vertebrate remains that have been amplified and sequenced using cyt b primers may be based on two techniques: BLAST searches (http://www.ncbi.nlm.nih.gov/BLAST/) or phylogenetic tree generation (Branicki et al. 2003, Honeycutt et al. 1995, Irwin et al. 1991, Parson et al. 2000). Both methods compare unknown and known sequences to determine the degree of divergence. BLAST is an internet-based program that compares an unknown sequence to known sequences and attempts to find the best matches. A non-redundant BLAST search, which filters out

identical sequences so these matches are not included, is performed and results are organized as a list of the top 100 comparisons ("hits"), arranged by degree of similarity. Included in the list are the species of origin, the gene identified, and information about sequence similarity. These include the 'bit score,' which is a value that indicates how similar two sequences are based on a pairwise comparison. The bit score, which is adjusted to take into account any gaps in the sequence alignment, increases with the similarity of the sequences and is used to calculate the 'e-value', which measures the likelihood of the sequence similarity being a result of chance as opposed to being a "real" match (Altschul et al. 1990, Hall 2001). E-values range between 0.0 and 1.0 with 0.0 corresponding to an exact match, therefore, the lower the number, the more confident one can be in a match.

Phylogenetic tree generation uses specific algorithms to compare sequences and generate the most likely evolutionary arrangement of a given set of species based on differences among the compared sequences. One requirement of tree generation is correct sequence alignment. Two programs that can be used for sequence alignment are Sequencher (by Genecodes) and MacClade (Maddison and Maddison 2000), but only Sequencher allows for the visualization of electropherograms for base editing. Edited and aligned sequences can be exported out of Sequencher in a compatible format for viewing in MacClade, where they are translated into amino acid codons. This can facilitate a more accurate alignment of the sequences because any gaps in the nucleotide sequences are adjusted based on the proper protein alignment. The realigned nucleotide sequences are transported into the Phylogenetic Analysis Using Parsimony program (PAUP), which presents several user-defined options for tree generation (Swofford

1998). One can choose what algorithm or method to use, whether to root the tree, and whether to perform a bootstrap evaluation after the tree has been generated. Trees can be generated using either tree-searching or distance-based methods, the former having higher discrimination capabilities but requiring more time, sometimes hours to days depending on the search (Hall 2001, Huelsenbeck et al. 1995, Maddison 2000, Takahashi et al. 2000). Therefore, in the interests of time and in consideration of the overall goals of this validation study, the distance-based neighbor joining method was chosen. This method begins with an unresolved (unorganized) group of sequences and gradually builds a single tree by pairing each sequence with another sequence such that the smallest sum of branch lengths is achieved (reviewed by Hall 2001). The neighbor joining method is algorithmic and determines relationships based on calculation of distances (number of sequence differences) to each branchpoint or node. A separate algorithm, Jukes-Cantor, is used to calculate these distances. Jukes-Cantor uses the minimum number of differences or minimum evolution principle, which is based on the concept that the end product would have been produced using the least number of nucleotide base changes (Takahashi and Nei 2000). The neighbor-joining method using Jukes Cantor is able to generate trees with at least 90% accuracy depending on the lengths of the branches (Kumar and Gadagkar 2000), though tree-searching methods can potentially be applied to the data for further discrimination capabilities (Hillis et al. 1994, Huelsenbeck 1995, Takahashi and Nei 2000, reviewed by Hall 2001).

In addition to choosing how to generate a tree, one must decide whether to generate unrooted or rooted trees. Unrooted trees branch out from a central point, thus giving a circular tree with no particular species acting as the beginning branchpoint.

Rooted trees, on the other hand, use a specific species as an outgroup from which all other clades (branch groupings) will stem; the chosen species is usually one that should only have a distant relationship to the potential species of unknown specimens and would not be grouped with any of the other species in the tree (Maddison 2000, reviewed by Hall 2001). For example, if generating a tree to determine the evolutionary relationships among all species of turtles, one might use a different reptile, such as a snake sequence, as the outgroup.

Finally, whereas BLAST uses e-values to determine confidence, PAUP allows for a bootstrap calculation after the tree is generated, which provides estimates of the confidence of the placement of each species in a tree by assigning individual bootstrap percentages to each branch of the tree. The bootstrap analysis chooses random trees out of all possible trees and conducts a resampling of a user-specified number of these trees (default =100) to determine how many have the same placement for the nodes or branchpoints. The more often a node appears in the same position among all of the trees, the higher its bootstrap value, or confidence level, will be.

Validation of Cytochrome b for AFDIL

Scientists at the Armed Forces DNA Identification Laboratory (AFDIL) chose to validate Parson et al.'s (2000) universal cyt b primers for species identification. AFDIL's primary mission is to aid the Central Identification Laboratory in Hawaii (CILHI) with the identification of human skeletal remains recovered from World War II, the Korean War, and the Vietnam conflicts. As the time span between wars and the recovery of remains increases so does the level of skeletal degradation. This can increase DNA amplification failure when dealing with small pieces of bone or highly degraded skeletal

remains that cannot be distinguished as human based on physical characteristics. Currently, AFDIL amplifies HVI (nt 15989-16410) and HVII (nt 15-389) regions with either four human specific primer sets for relatively intact mtDNA genomes or 8 human specific mini-primer sets for highly degraded or inhibited samples. When amplifications are successful, the product is sequenced, and the results are compared to known reference samples. However, valuable time and resources are wasted with additional troubleshooting efforts that attempt to control for inhibition, degradation, and low copy number when the extract is non-human. In these instances, a validated set of vertebrate specific primers that amplify a small, variable region among species, such as the cyt b gene described above, would be helpful for targeting causes of amplification failure.

The study described here builds upon a preliminary study conducted at AFDIL in 2000, during the course of which two George Washington University graduate students amplified 5 pg or more of vertebrate DNA using Kocher et al.'s (1989) PCR parameters and Parson et al.'s (2000) vertebrate cyt b primers (unpublished results). The current validation addressed several factors for use of the cyt b primers, including: (1) optimization of amplification conditions, which involved determination of the limit of detection and evaluation of effects of cycle number and annealing time increases, (2) vertebrate specificity of the primers, (3) sequence consistency among species when using the primers in terms of sequence length and quality, (4) species determination capabilities comparing two different methods, and (5) determination of mixture detection levels. The goal of the validation was to formulate a procedure for amplification and identification of DNA from skeletal remains for non-human/human classification using as little as 1 pg of input DNA for amplification.

MATERIALS AND METHODS

Genomic DNA

Whole bloodstains on FTA® cards were obtained from the College of Agriculture and Natural Resources of the University of Delaware for domestic cat (*Felis catus*), domestic dog (*Canis familiaris*), domestic sheep (*Ovis aries*), and domestic horse (*Equus caballus*). Genomic DNA extracts at known concentrations from alligator (*Alligator mississippiensis*), domestic cow (*Bos taurus*), gorilla (*Gorilla gorilla*), European rabbit (*Oryctolagus cuniculus*), and yeast (*Saccharomyces cerevisiae*), and an unknown concentration of brown kiwi DNA (*Apteryx australis mantelli*), were provided by Dr. Tom Parsons of AFDIL. All DNA extracts were stored at –20°C.

DNA extracts from bacteria, chicken, clam, fruit fly, lobster, marmoset, nematode, pig, and sea urchin (specific species unknown) were purchased from BIOS Laboratories at a concentration of 50 ng/µl and were stored at 4°C. Human genomic DNA from an AFDIL scientist [DAL] was organically extracted, quantified, diluted to 20 pg/µl, and stored at –20°C. This was used as the human positive control for all amplification procedures.

Cytochrome b Primers Synthesis

Cytochrome b primer sequences were identical to those used by Parson et al. (2000) and were: Cytb F (forward) 5'-CCATCCAACATCTCAGCATGATGAAA-3' and Cytb R (reverse) 5'-CCCCTCAGAATGATATTTGTCCTCA-3'. These primers are vertebrate specific and amplify an approximately 307 base pair segment from the 5' end of cytochrome b (Branicki et al. 2003, Irwin et al. 1989, Kocher et al. 1989, Parson et al. 2000). Synthesis was performed at AFDIL using the column-based phosphoramidite

method (Caruthers et al. 1983). Synthesized primers were removed from the synthesis column by the addition of 15μM ammonium hydroxide and collected into 2 ml collection vials. The collection vials were then placed in a 55°C oven for 8 hours to cleave protecting groups. This solution was distributed into eight 1.7 ml microcentrifuge tubes and dried under vacuum at 50°C for approximately 75 minutes. The primers were reconstituted by adding 300 μl of 10mM Tris, 0.1mM EDTA (TLE) to the first tube, pipetting to resuspend the DNA, and transferring the solution to subsequent tubes. For quantification, the primers were diluted 1:500 in TLE, and an A260 reading was taken using a spectrophotometer. The primers were diluted to 10 μM and distributed into 1.7 ml microcentrifuge tubes for storage at –20°C.

Amplification Optimization

To test for the presence of contaminating DNA that may have been introduced during primer synthesis, 50 μl amplification reactions were set up in 0.2 mL eppendorf tubes following the AFDIL Quality Control protocol: negative control 1, negative control 2, negative control 3, positive A (10pg), positive B (10pg), negative control 4, negative control 5. The PCR master mix contained 5 μl of GeneAmp 10X PCR Buffer (500mM KCl, 100mM Tris-HCl, pH 8.3; 1.5mM $MgCl_2$ and 0.01% (w/v) gelatin), 4 μl of 2.5mM dNTPs, 2 μl of 0.625 μg/μl BSA, 2 μl each of 10 μM forward and reverse primers (cytb F and cytb R), 2.5 μl of 5 U/μl AmpliTaq Gold DNA polymerase, and sterile dH_2O to a final 40 μl volume. The buffer, dNTPs, BSA, and water were added to the master mix first, and the solution was sterilized by U/V irradiation for 20 minutes. The remaining reagents (primers and AmpliTaq Gold) were added, and 40 μl of the master mix were transferred to each reaction tube followed either by 10 μl of 1 pg/μl DAL DNA for the

positive or 10 µl water for the negative reactions. Thirty cycle amplifications were initially performed using two different PCR programs (cyto 1 and cyto 2) in a Perkin Elmer 9700 Thermal Cycler utilizing the 9600 ramp speed (Table 1).

Amplification results were evaluated using a 2% agarose gel [1.2 g agarose and 60 ml 1X TBE Buffer (89mM Tris HCl, pH 8.3; 89mM boric acid; 2mM EDTA)] containing 3 µl of 5 mg/ml ethidium bromide. Five microliters of each reaction were added to 1 µl of 10X agarose gel loading buffer (50% glycerol, 1.5 mM bromophenol blue, 100 mM EDTA) and loaded onto the gel between two 123-bp ladders. The gel was electrophoresed at 160–170 V for approximately 12 minutes, visualized on an ultraviolet transilluminator, and photographed. Amplicons were evaluated for band intensity and for the correct size by comparing them to the 123-bp ladder fragments.

The sensitivity of the amplification was evaluated at 100 pg, 10 pg, and 1 pg of genomic control DNA [DAL (20 pg/µl stock solution)] and included a negative control as the first and last amplification sample. The stock DAL was diluted for the 10 pg and 1 pg reactions such that 5 µl of the dilution were added to each reaction. Amplifications were performed first using both cyto 1 and cyto 2 programs at 38 cycles, second with cyto 2 at 38 cycles with 10 seconds added to the annealing time, and third with cyto 2 using 42 cycles. During the 38 cycle amplifications, a portion of HVI (nucleotides 16190–16410 amplified by primer set 2) from DAL was used as a positive amplification control. The amplification reagent volumes and amplicon visualization were as described above. All subsequent amplifications were performed using the cyto 2-42 cycle program. The 1 pg, 10 pg, and 100 pg amplification products were each purified and sequenced as described in the sequencing section below to demonstrate that human DNA was amplified.

Table 1. Cyto 1 versus cyto 2. The cyto 1 and cyto 2 programs differ in the times designated for each of the cycle steps: denaturation, annealing, and extension. This table gives a side-by-side comparison of the differences between the programs.

	A. cyto 1 parameters	**B. cyto 2 parameters**
Initial denaturation:	96°C - 10 minutes	95°C - 10 minutes
30 cycles of:		
denaturation	94°C - 1 minute	94°C - 30 seconds
annealing	50°C - 1 minute	50°C - 45 seconds
extension	72°C - 1 minute	72°C - 45 seconds
Final extension	72°C - 7 minutes	72°C - 7 minutes
Soak	4°C - ∞	4°C - ∞

Database Development

A database of 94 vertebrate cyt b sequences were compiled from GenBank via the NCBI website (Table 2), including the 14 species for which DNA was amplified and sequenced during the course of this validation. All sequences were copied into Sequencher 4.1.1b and aligned. The aligned sequences were exported as a Nexus file for comparison using MacClade software, and phylogenetic trees were generated using PAUP software.

Chelex Extraction

Three 1/8" diameter punches from the FTA® card of each species (domestic cat, domestic dog, domestic sheep, domestic horse) were deposited into 1.7 ml microfuge tubes containing 1 ml U/V irradiated, de-ionized water. The FTA® cards were then vacuum-sealed in envelopes containing desiccant and stored in a -20°C freezer. Samples were vortexed and allowed to incubate for one hour. The samples were centrifuged for 3 minutes at 15,000 rpm, and all but 30 µl of the supernatant was discarded; then 170 µl of a 5% Chelex® solution (w/v) were added to each sample. The samples were incubated for one hour at 55°C, vortexed for 10 seconds, incubated in a boiling water bath for 8 minutes, and vortexed for 10 seconds. The Chelex® resin and blood punch were pelleted at 15,000 rpm for 3 minutes, and the samples were stored at 4°C.

Species Differentiation

DNAs from alligator, chicken, domestic cow, gorilla, marmoset, house mouse, pig, and European rabbit were serially diluted so that stocks yielded a total of either 0.1 pg/µl or 1 pg/µl, respectively. Ten µl of each diluted DNA were amplified using the cyt b primers as described above. Brown kiwi was amplified using 5 µl of the original

Table 2. Database Classification Table. Includes the class, order, family, common and species names, and GenBank Accession numbers of the 95 GenBank sequences. Common names in bold designate species that were also tested during the course of this study.

Common Name	***Scientific Name***	**Class**	**Order**	**Family**	**GenBank Accession #**
Trout, Rainbow	*Oncorhynchus mykiss*	Actinopterygii	Salmoniformes	Salmonidae	L29771
Frog, Bull	*Rana catesbeiana*	Amphibia	Anura	Ranidae	AF205089
Frog, Wrinkled	*Rana rugosa*	Amphibia	Anura	Ranidae	AF205092
Salamander, Luschan's	*Mertensiella luschani*	Amphibia	Caudata	Salamandridae	AF154053
Duck, Wandering Whistling	*Dendrocygna arcuata*	Aves	Anseriformes	Dendrocygnidae	AF082061
Hawk, Sharp-shinned	*Accipiter striatus*	Aves	Ciconiiformes	Accipitridae	U83305
Heron, Grey	*Ardea cinerea*	Aves	Ciconiiformes	Ardeidae	AF375962
Stork, Oriental White	*Ciconia boyciana*	Aves	Ciconiiformes	Ciconiidae	AB026193
Vulture, Greater Yellow-headed	*Cathartes melambrotus*	Aves	Ciconiiformes	Ciconiidae	AF494340
Falcon, Peregrine	*Falco peregrinus*	Aves	Ciconiiformes	Falconidae	U83307
Chicken	*Gallus gallus*	Aves	Galliformes	Phasianidae	AY170102
Pheasant, Edward's	*Lophura edwardsi*	Aves	Galliformes	Phasianidae	AF534557
Quail, Gambel's	*Lophortyx gambelii*	Aves	Galliformes	Phasianidae	L08382
Turkey, Wild	*Meleagris gallopavo*	Aves	Galliformes	Phasianidae	L08381
Crane, Japanese	*Grus japonensis*	Aves	Gruiformes	Gruidae	U11063
Crow, American	*Corvus brachyrynchos*	Aves	Passeriformes	Corvoidea	AY030112
Robin, American	*Turdus migratorius*	Aves	Passeriformes	Turdidae	AF197835
Thrush, Eye-browed	*Turdus obscurus*	Aves	Passeriformes	Turdidae	AY049484
Owl, Little	*Asio flammeus*	Aves	Strigiformes	Strigidae	U89171
Kiwi, Brown	*Apteryx australis*	Aves	Struthioformes	Apterygidae	U76050
Ostrich	*Struthio camelus*	Aves	Struthioformes	Struthionidae	U76055
Shark, Small-tailed	*Carcharhinus porosus*	Chondrichthyes	Carcharhiniformes	Carcharhinidae	L08033
Shark, Tiger	*Galeocerdo cuvier*	Chondrichthyes	Carcharhiniformes	Carcharhinidae	L08034
Antelope, Sable	*Hippotragus niger*	Mammalia	Artiodactyla	Bovidae	AF036285
Bison, American	*Bison bison*	Mammalia	Artiodactyla	Bovidae	AF036273

Cow, Domestic	*Bos Taurus*	Mammalia	Artiodactyla	Bovidae	D34635
Goat, Domestic	*Capra hircus*	Mammalia	Artiodactyla	Bovidae	AB004070
Jentink's Duiker	*Cephalophus jentinki*	Mammalia	Artiodactyla	Bovidae	AF153888
Sheep, Domestic	*Ovis aries*	Mammalia	Artiodactyla	Bovidae	AB006800
Llama	*Lama glama*	Mammalia	Artiodactyla	Camelidae	U06429
Deer, Mule	*Odocoileus hemionus*	Mammalia	Artiodactyla	Cervidae	AF091630
Deer, Roe	*Capreolus capreolus*	Mammalia	Artiodactyla	Cervidae	Y14951
Giraffe	*Giraffa camelopardalis*	Mammalia	Artiodactyla	Giraffidae	X56287
Hippopotamus	*Hippotomaus amphibious*	Mammalia	Artiodactyla	Hippopotamidae	U07565
Boar, Wild	*Sus scrofa*	Mammalia	Artiodactyla	Suidae	Z50089
Dog	*Canis familiaris*	Mammalia	Carnivora	Canidae	X94920
Fox, Red	*Vulpes vulpes*	Mammalia	Carnivora	Canidae	X94929
Wolf, Gray	*Canis lupus*	Mammalia	Carnivora	Canidae	AY170103
Cat, Domestic	*Felis catus*	Mammalia	Carnivora	Felidae	X82296
Cat, Wild	*Felis silvestris*	Mammalia	Carnivora	Felidae	AY170102
Lion	*Panthera leo*	Mammalia	Carnivora	Felidae	X82300
Tiger	*Panthera tigris tigris*	Mammalia	Carnivora	Felidae	X82301
Hyena, Spotted	*Crocuta crocuta*	Mammalia	Carnivora	Hyaenidae	AY170114
Badger, Eurasian	*Meles meles*	Mammalia	Carnivora	Mustelidae	X94922
Marten, American	*Martes americana*	Mammalia	Carnivora	Mustelidae	AF154968
Mink, American	*Mustela vison*	Mammalia	Carnivora	Mustelidae	AB026109
Polecat	*Mustela putorius*	Mammalia	Carnivora	Mustelidae	AB026107
Sable	*Martes zibellina*	Mammalia	Carnivora	Mustelidae	AB012358
Weasel, Long-tailed	*Mustela frenata*	Mammalia	Carnivora	Mustelidae	AF498153
Bear, Asiatic Black	*Ursus thibetanus*	Mammalia	Carnivora	Ursidae	AB020910
Bear, Brown	*Ursus arctos*	Mammalia	Carnivora	Ursidae	AB020909

Bear, Polar	*Ursus maritimus*	Mammalia	Carnivora	Ursidae	U18898
Panda, Giant	*Ailuropoda melanoleuca*	Mammalia	Carnivora	Ursidae	U23552
Dolphin, Gray	*Sotalia fluviatilis*	Mammalia	Cetacea	Delphinidae	AF304067
Whale, Sperm	*Physeter catodon*	Mammalia	Cetacea	Physeteridae	AF304073
Bat, Fruit	*Artibeus phaeotis*	Mammalia	Chiroptera	Phyllostomidae	U66514
Possum, Black "Four-eyed"	*Philander mcilhennyi*	Mammalia	Didelphimorphia	Didelphiae	U34680
Shrew, Eurasian common	*Sorex araneus*	Mammalia	Insectivora	Soricidae	AJ000416
Hare	*Lepus capensis*	Mammalia	Lagomorpha	Leporidae	AJ279417
Rabbit, European	*Oryctolagus cuniculus*	Mammalia	Lagomorpha	Leporidae	U07566
Pika, Daurian	*Ochotona daurica daurica*	Mammalia	Lagomorpha	Ochotonidae	AF273011
Donkey, Domestic	*Equus asinus*	Mammalia	Perissodactyla	Equidae	X97337
Horse, Domestic	*Equus caballus*	Mammalia	Perissodactyla	Equidae	D32190
Zebra, Grevy's	*Equus grevyi*	Mammalia	Perissodactyla	Equidae	X56282
Rhino, Black	*Diceros bicornis*	Mammalia	Perissodactyla	Rhinocerotidae	X56283
Rhino, Sumatran	*Dicerorhinus sumatrensis*	Mammalia	Perissodactyla	Rhinocerotidae	AJ245723
Marmoset	*Callithrix jacchus*	Mammalia	Primates	Callitrichidae	AF295586
Howler Monkey, Black	*Alouatta caraya*	Mammalia	Primates	Cebidae	AF289519
Monkey, Spider	*Ateles geoffroyi panamensis*	Mammalia	Primates	Cebidae	AY065903
Uakari, Black	*Cacajao melanocephalus*	Mammalia	Primates	Cebidae	AF524891
Baboon	*Papio hamadryas*	Mammalia	Primates	Cercopithecidae	Y16590
Colobus, Angolan	*Colobus angolensis*	Mammalia	Primates	Cercopithecidae	AF295583
Langur, Douc	*Pygathrix nemaeus*	Mammalia	Primates	Cercopithecidae	AF295582
Macaque, Lion Tail	*Macaca silenus*	Mammalia	Primates	Cercopithecidae	AF350404
Monkey, Rhesus	*Macaca mulatto*	Mammalia	Primates	Cercopithecidae	U38272
Human	*Homo sapiens*	Mammalia	Primates	Hominids	U09500
Lemur, Ring-tailed	*Lemur catta*	Mammalia	Primates	Lemuridae	U53575

Chimpanzee	*Pan troglodytes*	Mammalia	Primates	Pongidae	X93338
Gorilla	*Gorilla gorilla*	Mammalia	Primates	Pongidae	D38114
Orangutan	*Pongo pygmaeus*	Mammalia	Primates	Pongidae	U38274
Elephant, African	*Loxodonta africana*	Mammalia	Proboscidea	Elephantidae	D84152
Elephant, Asian	*Elephas maximus*	Mammalia	Proboscidea	Elephantidae	AB002412
Beaver, European	*Castor fiber*	Mammalia	Rodentia	Castoridae	AJ389529
Hamster	*Cricetulus griseus*	Mammalia	Rodentia	Muridae	AB033693
Mouse, House	*Mus musculus*	Mammalia	Rodentia	Muridae	AY057807
Muskrat	*Ondatra zibethicus*	Mammalia	Rodentia	Muridae	AF119277
Rat, Brown	*Rattus norvegicus*	Mammalia	Rodentia	Muridae	AB033713
Squirrel	*Spermophilus variegates*	Mammalia	Rodentia	Sciuridae	AF157854
Dugong	*Dugong Dugong*	Mammalia	Sirenia	Dugongidae	U07564
Alligator	*Alligator mississippiensis*	Reptilia	Crocodylia	Alligatoridae	Y13113
Caiman	*Caiman crocodylus*	Reptilia	Crocodylia	Alligatoridae	AJ404872
Boa	*Boa constrictor*	Reptilia	Squamata	Boidae	AF471036
King Cobra	*Ophiphagus hannah*	Reptilia	Squamata	Elapidae	AF217842
Iguana	*Iguana iguana*	Reptilia	Squamata	Iguanidae	U88954
Turtle, Box	*Terrapene carolina*	Reptilia	Testudinata	Emydidae	AF258871

extract since the concentration was not specified. The Chelex®-extracted DNAs were not quantified. These samples were diluted 1:1000 (1 µl of Chelex® product was added to 999 µl of water), and 10 µl of this dilution were amplified under the same conditions as the other vertebrate DNAs except that 0.5 µl of AmpliTaq Gold DNA polymerase were used. After two months storage, the ovine, canine, and equine extracts did not generate detectable amplicons using the original 1:1000 dilution or a 1:500 dilution. Amplifications were repeated using 1:10 dilutions with 2 µl of DNA and 1 µl AmpliTaq Gold DNA polymerase to determine if the DNA was degrading during storage. In addition to the human positive control, an invertebrate control (yeast or nematode) that was not expected to amplify was included with each set of reactions. Finally, a series of invertebrates (bacteria, clam, fruit fly, lobster, nematode, sea urchin, and yeast) were amplified in 50 µl reactions using 5 µl of DNA at 20 pg/µl. As described above, all products were visualized by agarose gel electrophoresis.

The amplification products were purified using Centricon-100® spin filtration units as follows: (1) Two ml of sterile dH_2O and the PCR product (45 µl) were added to the column, which was centrifuged at 1000 x g for 20 minutes; (2) An additional 2 ml of sterile water were added, and the centrifugation was repeated as in step 1; (3) The reservoir was flipped and centrifuged at 1000 x g for two minutes to recover the purified amplicon; (4) All samples were brought to a final volume of 50 µl with sterile dH_2O and stored at 4°C.

Sequence reactions were performed using an ABI Dye Terminator Cycle Sequencing Ready Reaction Kit containing AmpliTaq DNA Polymerase (BigDye version 1.0). Reactions were set-up in 96 well optical plates on ice and included 2 – 8 µl of DNA

(depending on the intensity of the band on the agarose gel), 1 µl of 10 µM primer, 8 µl of Big Dye version 1.0, and sterile water to 20 µl. The wells were covered with strip caps, vortexed, and subjected to 25 cycles of (96°C, 15 sec.; 50°C, 5 sec.; 60°C, 2 min.). Sequencing products were purified in a Performa® DTR 96-well standard purification plate according to the manufacturer's protocol (EDGE Biosystems). The purified samples were transferred to a 96 well optical plate and dried in a heated vacuum concentrator for 50 – 60 minutes then sealed and stored at –20°C.

Sequencing products were reconstituted by adding 10 µl of HiDi-formamide to each well. Optical plates were covered with a 96 well septa, and the plates were vortexed to mix and centrifuged for 1 minute. Each of the optical plates was placed into a 3100 plate base with retainer and positioned on the autosampler deck (two plates per run). Sample sheets were created using the 3100 Data Collection software with the parameters: Dye Set E, DT3100POP6(BD)v2.mob mobility file, the RapidSeq36 POP6Module1 run module, and the BC-3100RR SeqOffFtOff.saz analysis module. Sequencing samples were electrokinetically injected for 15 seconds at 3 kV and electrophoresed on a 36 cm array for 40 minutes at 15 kV and 55°C. The data files were extracted automatically to the server and analyzed using Sequence Analysis NT version 3.7 or higher. All files except the amplification controls and reagent blanks were analyzed with "PCR stop setting" used to end all sample sequences after a run of 10 uncalled nucleotides (N). The amplification controls and reagent blanks were analyzed using the default settings, which analyze the entire sequence files. Electropherograms were printed and data files analyzed using Sequencher. The forward and reverse sequences for each sample were aligned automatically using the parameters: assembly algorithm = clean data; minimum match

percentage = 80%; and minimum overlap = 20 base pairs. The aligned sequences were visually evaluated for peak height definition and amplitude within the call region (the amplified segment between the forward and reverse primer sequences). Only sequences with a peak height of at least 25 RFU's were considered acceptable as this is the cutoff for samples analyzed on the ABI 3100 Genetic Analyzer at AFDIL. Ambiguous peaks that could not be resolved by eye as well as any heteroplasmic peaks were designated as N's, and the consensus sequences were then saved as text files.

BLAST Comparison and Phylogenetic Analysis

The text files were imported into BLAST, and a non-redundant nucleotide-nucleotide BLAST search (blastn) was conducted for each sequence to determine the closest species match. BLAST results were evaluated based on the species and e-value of the top matches, or "hits." The determined consensus sequence for each species was also aligned with the corresponding GenBank reference species sequence to evaluate the exact number of differences between the experimental and reference sequences to determine if there was a correlation between the number of differences and the resultant e-value.

Sample consensus sequences were copied into a Sequencher file containing the 94 GenBank vertebrate species sequences (See Database Development: Table 2). All known and experimental sequences were aligned and exported as a Nexus file. Phylogenetic comparisons were made using MacClade and PAUP. MacClade was used to translate the aligned sequences into proteins and to initiate alignment based on codon sequences. The realigned set of sequences was then imported into PAUP for phylogenetic tree generation. Rooted trees were created and bootstrap analysis conducted based on distance using the neighbor-joining method with the Jukes-Cantor algorithm (Efron et al.

1996; Hall 2001). The bootstrap calculations were used as indications of the confidence of the tree placement of each species. The rainbow trout (*Onchorynchus mykiss*) sequence was chosen as the outgroup (or root) for all trees.

Three different evolutionary trees were generated using PAUP. The first tree used only the ninety-four GenBank database sequences as a test to determine whether species would be grouped accurately based on class and family relationships. The goal of generating the second tree was to test whether the experimental alligator, cow, pig, cat, kiwi, marmoset, human, gorilla, chicken, and rabbit sequences were aligned correctly with their respective GenBank database sequences for observation of placement. For the final tree, the eleven GenBank database species that matched those that were tested during the course of this validation were removed from the set of sequences that had been used to generate the second tree. The purpose of this tree was to determine how similar the family placement for the experimental sequences would be as compared to the placement for the same GenBank species.

Mixture Analysis.

Invertebrate:vertebrate mixtures were prepared using yeast:DAL, sea urchin:DAL, or lobster:DAL (Table 3A). Non-human vertebrate:human mixtures were prepared as alligator:DAL, chicken:DAL, or gorilla:DAL (Table 3B). Amplification, purification, and sequencing were performed as described in the amplification optimization and species differentiation sections. The total input DNA for the mixtures was 1 pg using varying combinations of 0.1 pg/µl solutions of each species.

Mixture sequences were assessed in Sequencher for separation of a major sequence from a minor sequence using the automatic assembly option, which

mechanically aligns the major mixture component with the GenBank database sequence for one or the other species comprising the mixture. Automatic assembly was considered successful if the major and minor components were resolved enough to be able to clearly distinguish the major sequence, meaning that the primary and secondary sequences could be clearly separated. The sequence for each mixture in the series was labeled as either the non-human component, human, or inconclusive based on the ability to determine a primary and secondary contributor to the mixture.

Table 3. Mixture Dilution Table. A) Invertebrate:human – For all reactions (except 100:0), 1 µl of a 1:20 dilution (1 pg/µl) of human DNA was added. B) Non-human vertebrate:human – Dilutions were mixed such that 1 pg total vertebrate DNA was added to each reaction.

Table 3A:	
Invertebrate:Vertebrate(Human)	**Invertebrate Dilutions**
100:0	100pg/5µl (5 µl)
100:1	100pg/5µl (5 µl)
50:1	100pg/5µl (2.5 µl)
10:1	10pg/5µl (5 µl)
5:1	10pg/5µl (2.5 µl)
1:1	10pg/5µl (0.5 µl)
0:1	

Table 3B:		
Vertebrate (non-human):Vertebrate (Human)	**Non-human volume (of 0.1pg/µl dilution)**	**Human volume (of 0.1 pg/µl dilution)**
10:1	10µl	1µl
9:1	9µl	1µl
3:2	6µl	4µl
1:1	5µl	5µl
2:3	4µl	6µl
1:9	1µl	9µl
1:10	1µl	10µl

RESULTS

Amplification Parameters

Ten pg of human genomic amplification controls (Figs. 2A and B; lanes 2–8). When increased to 38 cycles, faint bands migrated at the predicted ~ 350 bp (based on the 123 bp ladder) for both the cyto 1 and cyto 2 parameters when 10 pg and 100 pg of genomic DNA were amplified (Figs. 3A and 3B; lanes 2–5 and lanes 3–6, respectively), and all negative amplification controls were clean (Fig. 3A and 3B; Lanes 1 and 9 and lanes 2 and 10, respectively). However, the 10 pg amplicon bands were more intense for the cyto 2 than for the corresponding cyto 1 samples, and no bands were visible for the 1 pg samples amplified using the cyto 1 program though one of the 1 pg specimens yielded visible product with the cyto 2 program (Fig. 3A and 3B; Lanes 2–3 and 3–4, respectively).

Based on the results described above, the cyto 2 parameters were further optimized by amplifying 1, 10 and 100 pg of human genomic DNA at 38 cycles with 10 seconds added to the annealing time or at 42 cycles. Little to no difference in amplification efficiency was observed for amplification at 38 cycles with 10 seconds annealing time versus the original 38 cycle program (compare Fig. 3B; lanes 3–6 with Fig. 3C; lanes 2–5). The 1 pg samples still produced no observable band with the increase in annealing time, and the 10 pg bands and 100 pg bands were of the same intensity as for 38 cycles. In contrast, at 42 cycles the 1 pg bands were visible, and all bands were of greater intensity than for 38 cycles or 38 cycles plus 10 seconds annealing time (Fig. 3D; lanes 3–4). In all instances, the PCR negative amplification controls were clean and HVI positive controls were as expected. These results generated an optimal

amplification protocol for the cyto b primers of 1 cycle of 95°C for 10 min followed by 42 cycles of 95°C for 30 sec, 50° for 45 sec, and 72° for 45 sec; and a 7 minute final extension.

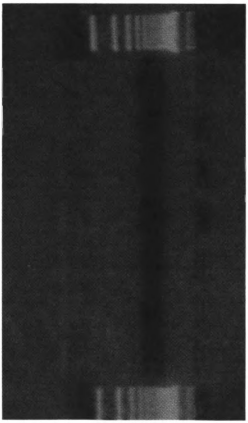
Using the optimized amplification parameters, a newly-synthesized lot of cyto b primers was evaluated for contamination and sensitivity using the cyto 2 - 42 cycle program. Results shown in Fig. 4 demonstrate that no detectable bands were present in the five negative amplification controls, but amplification products were detected for the 10 pg positive control samples (Fig. 4; Lanes 5–6), thus confirming that this lot of primers was contaminant free. The primers were then used to amplify 1, 10 and 100 pg of human genomic DNA, and the resulting 1 pg products were purified and sequenced as described in the amplification optimization section of the Materials and Methods.

A 307 bp region, not including the primer binding region, was confirmed by the forward and reverse sequences. There was one difference (a T → C transition at position 274) between the human GenBank known sequence and the human positive (DAL) sequence (Fig. 5). The top match from a BLAST search of the confirmed human positive sequence was the partial mitochondrial genome of a cloned human mtDNA (GenBank Accession # AF465976.1) with an e-value of e^{-171}.

Figure 2. Cytochrome b Primer Optimization at 30 Amplification Cycles. Ten pg of human control DNA were amplified simultaneously using either the cyto 1 or cyto 2 programs. A) Amplifications using the cyto 1 program with lane numbers and samples designated at the top. B) Amplifications using the cyto 2 program with lane numbers and samples designated at the top.

Lane 1-123 bp ladder
Lane 2-Negative 1
Lane 3-Negative 2
Lane 4-Negative 3
Lane 5-Positive A
Lane 6-Positive B
Lane 7-Negative 4
Lane 8-Negative 5
Lane 9-123 bp ladder

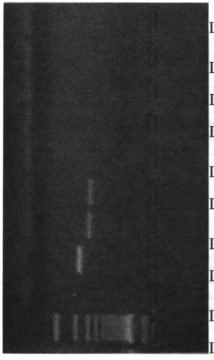
Fig. 2A. cyto 1–30 cycles



Lane 1-123 bp ladder
Lane 2-Negative 1
Lane 3-Negative 2
Lane 4-Negative 3
Lane 5-Positive A
Lane 6-Positive B
Lane 7-Negative 4
Lane 8-Negative 5
Lane 9-123 bp ladder

Fig. 2B. cyto 2–30 cycles

Fig. 3. Optimization Results. The cyto 1 and cyto 2 cycling parameters using 1, 10, and 100 pg of genomic DNA were re-evaluated using 38 cycles. Further optimization involved amplification of 1, 10, and 100 pg using at either 38 cycles with ten seconds added to the annealing time or at 42 cycles. Positive A and B indicate different aliquots of the human genomic DNA. Lane numbers and samples designations are at the top of each figure. A) cyto 1–38 amplification B) cyto 2–38 amplification cycles. C) cyto 2–38 cycles with 10-sec on the annealing time. D) cyto 2–42 cycles.



Lane 1-Negative 1

Lane 2-1pg Positive A

Lane 3-1pg Positive B

Lane 4-10pg Positive A

Lane 5-10pg Positive B

Lane 6-100 pg Positive A

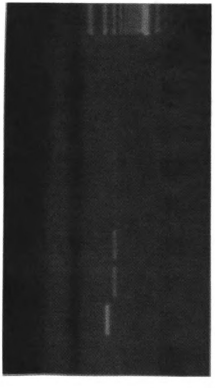
Lane 7-100 pg Positive B

Lane 8-Positive (HV1 region)

Lane 9-Negative 2

Lane 10-123 bp ladder

Fig. 3A. cyto 1–38 cycles



Lane 10-123 bp ladder

Lane 2-Negative 1

Lane 3-1pg Positive A

Lane 4-1pg Positive B

Lane 5-10pg Positive A

Lane 6-10pg Positive B

Lane 7-100 pg Positive A

Lane 8-100 pg Positive B

Lane 9-Positive (HV1 region)

Lane 10-Negative 2

Fig. 3B. cyto 2 – 38 cycles

Lane 1-Negative 1

Lane 2-1pg Positive A

Lane 3-1pg Positive B

Lane 4-10pg Positive A

Lane 5-10pg Positive B

Lane 6-100 pg Positive A

Lane 7-100 pg Positive B

Lane 8-Positive (HV1 region)

Lane 9-Negative 2

Lane 10-123 bp ladder

Fig. 3C. cyto 2–38 cycles with 10 seconds on the annealing time

Lane 1-123 bp ladder

Lane 2-Negative 1

Lane 3-1pg Positive A

Lane 4-1pg Positive B

Lane5-10pg Positive A

Lane 6-10pg Positive B

Lane 7-100 pg Positive A

Lane 8-100 pg Positive B

Lane 9-Negative 2

Lane 10-123 bp ladder

Fig. 3D. cyto 2–42 cycles

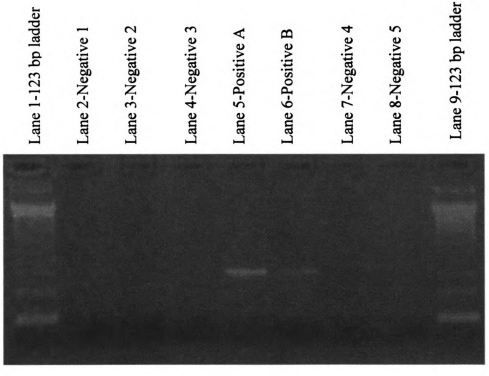


Fig. 4. Primer Quality Control. A newly synthesized lot of cyt b primers was evaluated for contamination using the cyto 2–42 program. The results demonstrated that all negative amplification control samples (lanes 2, 3, 4, 7, and 8) were clear.

Fig. 5. One pg Human DNA Sequence Alignment Results. The 1 pg sequencing results were aligned with the human cyt b reference. Differences are denoted with black dots below the consensus. The human GenBank reference sequence is the top sequence labeled Homo sapiens. The DAL forward sequence is labeled P1A1_CYF, and the reverse sequence is labeled P1A1_CYR. Numbering is based on the starting base of each individual sequence so the P1A1_CYF sequence will be numbered as one less than the other two sequences because one base is missing at the beginning.

```
Homo sapiens      #1  CTTCGGCTCA CTCCTTGGCG CCTGCCTGAT CCTCCAAATC ACCACAGGAC
P1A.1_CYF148... >#1>   TTCGGCTCA CTCCTTGGCG CCTGCCTGAT CCTCCAAATC ACCACAGGAC
P1A.1_CYR15173... #1  CTTCGGCTCA CTCCTTGGCG CCTGCCTGAT CCTCCAAATC ACCACAGGAC
                      ..........................................................
                    #1  CTTCGGCTCA CTCCTTGGCG CCTGCCTGAT CCTCCAAATC ACCACAGGAC

Homo sapiens     #51  TATTCCTAGC CATGCACTAC TCACCAGACG CCTCAACCGC CTTTTCATCA
P1A.1_CYF1481... #50  TATTCCTAGC CATGCACTAC TCACCAGACG CCTCAACCGC CTTTTCATCA
P1A.1_CYR1517... #51  TATTCCTAGC CATGCACTAC TCACCAGACG CCTCAACCGC CTTTTCATCA
                      ..........................................................
                   #51  TATTCCTAGC CATGCACTAC TCACCAGACG CCTCAACCGC CTTTTCATCA

Homo sapiens    #101  ATCGCCCACA TCACTCGAGA CGTAAATTAT GGCTGAATCA TCCGCTACCT
P1A.1_CYF148... #100  ATCGCCCACA TCACTCGAGA CGTAAATTAT GGCTGAATCA TCCGCTACCT
P1A.1_CYR151... #101  ATCGCCCACA TCACTCGAGA CGTAAATTAT GGCTGAATCA TCCGCTACCT
                      ..........................................................
                   #101  ATCGCCCACA TCACTCGAGA CGTAAATTAT GGCTGAATCA TCCGCTACCT

Homo sapiens    #151  TCACGCCAAT GGCGCCTCAA TATTCTTTAT CTGCCTCTTC CTACACATCG
P1A.1_CYF148... #150  TCACGCCAAT GGCGCCTCAA TATTCTTTAT CTGCCTCTTC CTACACATCG
P1A.1_CYR151... #151  TCACGCCAAT GGCGCCTCAA TATTCTTTAT CTGCCTCTTC CTACACATCG
                      ..........................................................
                   #151  TCACGCCAAT GGCGCCTCAA TATTCTTTAT CTGCCTCTTC CTACACATCG

Homo sapiens    #201  GGCGAGGCCT ATATTACGGA TCATTTCTCT ACTCAGAAAC CTGAAACATC
P1A.1_CYF148... #200  GGCGAGGCCT ATATTACGGA TCATTTCTCT ACTCAGAAAC CTGAAACATC
P1A.1_CYR151... #201  GGCGAGGCCT ATATTACGGA TCATTTCTCT ACTCAGAAAC CTGAAACATC
                      ..........................................................
                   #201  GGCGAGGCCT ATATTACGGA TCATTTCTCT ACTCAGAAAC CTGAAACATC

Homo sapiens    #251  GGCATTATCC TCCTGCTTGC AACTATAGCA ACAGCCTTCA TAGGCTATGT
P1A.1_CYF148... #250  GGCATTATCC TCCTGCTTGC AACCATAGCA ACAGCCTTCA TAGGCTATGT
P1A.1_CYR151... #251  GGCATTATCC TCCTGCTTGC AACCATAGCA ACAGCCTTCA TAGGCTAT
                      ..........................................................
                   #251  GGCATTATCC TCCTGCTTGC AACCATAGCA ACAGCCTTCA TAGGCTATGT
                                                          •
Homo sapiens    #301  CCTCCCG
P1A.1_CYF148... #300  CCTCCCG
                      .......
                   #301  CCTCCCG
```

Fig. 5. (cont.)

Primer Specificity

The primers were assessed for their ability to amplify DNA from 14 different vertebrate species and inability to amplify DNA from seven different invertebrates. Results demonstrated that the cyt b primers did not amplify 100 pg of DNA from the seven invertebrate species (Fig. 6). Alternatively, the 1 pg and 10 pg organically-extracted vertebrate DNAs generated approximately a 350 bp fragment when compared to the 350 bp band of the 123 bp ladder (Fig. 7A–D; Lanes 3–6 and 8). Negative amplification and specificity controls were clean (Fig. 7A, B, and D; Lanes 2, 7, and 9; Fig. 7C; Lanes 2, 6, and 8). Comparison of the gel band intensities for all quantified species revealed that similar intensities were achieved for the 10 pg amplification products (Fig. 7A), as were the 1 pg specimens except for the gorilla (darker) and American alligator (lighter).

Similarly, the 1:1000 dilutions of Chelex®-extracted vertebrate DNAs produced detectable amplicons of the expected size and of similar intensities (Fig. 8A, Lanes 4–7). Again, the reagent blank and the amplification and specificity controls were clean (Fig. 8A; Lanes 2–3, 9–10). After two months storage at 4°C, the 1:10 dilutions generated detectable bands for all DNAs tested (Fig. 8B; Lanes 4–6). The amplification and specificity controls did not produce detectable products (Fig. 8B; Lanes 2, 8, and 9).

Species Sequencing Results

Sequencing was attempted for all invertebrate amplification product, and no detectable sequences were obtained for any of the invertebrate species. Non-contaminated sequences (single source) were generated for 11 of the 14 species tested (American alligator, kiwi, chicken, cat, cow, pig, rabbit, gorilla, human, marmoset, and

horse). All 1pg sequences displayed well-defined peaks between 200 and 500 relative fluorescence units (RFUs). The 10 pg sequences displayed RFUs approximately 10 fold higher in intensity. Nine of the eleven single source species were matched with a correct species with BLAST, corresponding to either the mitochondrial genome or the cyt b gene for the matching species with e-values ranging from e^{-171} to e^{-164} (Table 4). In addition, all nine sequences aligned with their respective GenBank reference sequences with no more than three nucleotide differences (Table 4). The remaining two species (marmoset and domestic horse) differed from their respective control sequences by over 50 bases. The top BLAST match for the marmoset was the Cotton-topped Tamarin (*Saguinus oedipus*) with an e-value of e^{-164}, and the actual marmoset sequence was 16th on the list of matches with an e-value of $2e^{-30}$. Similarly, the top BLAST result for the horse sequence, zebra (*Equus grevyi*), was inconsistent with the expected species. The zebra match showed an e-value of $3e^{-86}$.

The three remaining vertebrate species sequences (house mouse, domestic sheep, and domestic dog) exhibited evidence of contamination as indicated by the presence of two overlapping peaks at numerous positions. The sequencing results from the three contaminated species generated read lengths of 305 to 307 bp. Low level contamination, less than 10% of the major peak height, was observed for the house mouse, but the minor peak heights for the domestic sheep and domestic dog were at least 50% and at times equal to the major peak heights (Fig. 9 A–C). The major contributing sequences from the contaminated samples were determined and aligned with their appropriate GenBank control sequences and entered into BLAST. The top BLAST match for the house mouse was the house mouse cyt b gene with an e-value of e^{-171} with only one difference from the

Lane 1-123 bp ladder
Lane 2-Negative 1
Lane 3-Bacteria
Lane 4-Clam
Lane 5-Fruit Fly
Lane 6-Lobster
Lane 7-Nematode
Lane 8-Sea Urchin
Lane 9-Yeast
Lane 10 -Positive Control
Lanne 11-Negative 2
Lane 12-123 bp ladder

Fig. 6. Invertebrate Specificity Experiment. Agarose gel of invertebrate samples amplified using 100 pg of total input DNA. Products were visualized by ETBR gel electrophoresis.

Fig. 7. Vertebrate Sensitivity ETBR Agarose Gel Images. One and 10 pg of vertebrate DNA were amplified and visualized by ETBR agarose gel electrophoresis. **(A)** 10 pg amplification results for American alligator, European rabbit, and marmoset. Brown kiwi was of unknown concentration. **(B)** Amplification results for the 10 pg of chicken, cow, house mouse, and pig. **(C)** One pg amplification of American alligator, gorilla, and European rabbit. **(D)** One pg amplification results for chicken, cow, house mouse, and pig.

Fig. 7A.

Lane 1-123 bp ladder

Lane 2-Negative 1

Lane 3-American alligator

Lane 4-Brown kiwi

Lane 5-European rabbit

Lane 6-Marmoset

Lane 7-Invertebrate Control

Lane 8-Positive Control

Lane 9-Negative 2

Lane 10-123 bp ladder

Fig. 7B.

Lane 1-123 bp ladder

Lane 2-Negative 1

Lane 3-Chicken

Lane 4-Domestic cow

Lane 5-House mouse

Lane 6-Pig

Lane 7-Invertebrate Control

Lane 8-Positive Control

Lane 9-Negative 2

Lane 10-123 bp ladder

Fig. 7C.

Lane 1-123 bp ladder

Lane 2-Negative 1

Lane 3-American alligator

Lane 4-Gorilla

Lane 5-European rabbit

Lane 6-Invertebrate Control

Lane 7-Positive Control

Lane 8-Negative 2

Lane 9-123 bp ladder

Fig. 7D.

Lane 1-123 bp ladder

Lane 2-Negative 1

Lane 3-Chicken

Lane 4-Domestic cow

Lane 5-House mouse

Lane 6-Pig

Lane 7-Invertebrate Control

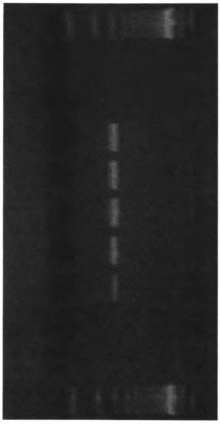
Lane 8-Positive Control

Lane 9-Negative 2

Lane 10-123 bp ladder

Fig. 8. Chelex® Extracted Vertebrate DNA Specificity. FTA™ bloodstain cards were Chelex® extracted, boiled, and 2 μl of a 1:1000 dilution of the extracted DNA amplified. **(A)** Agarose gel image for the 1:1000 dilution. **(B)** Agarose gel image of the 1:10 dilution of the Chelex® extracted samples.

Fig. 8A.



Lane 1-123 bp ladder

Lane 2-Negative 1

Lane 3-Reagent Blank

Lane 4-Domestic sheep

Lane 5-Domestic cat

Lane 6-Domesti dog

Lane 7-Domestic horse

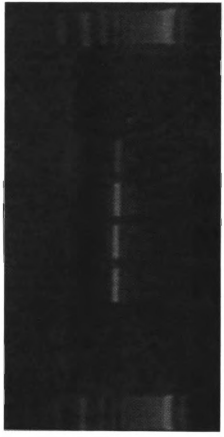
Lane 8-Positive Control

Lane 9-Invertebrate Control

Lane 10-Negative 2

Lane 11-123 bp ladder

Fig. 8B.



Lane 1-123 bp ladder

Lane 2-Negative 1

Lane 3-Reagent Blank

Lane 4-Domestic sheep

Lane 5-Domestic dog

Lane 6-Domestic horse

Lane 7-Positive Control

Lane 8-Invertebrate Control

Lane 9-Negative 2

Lane 10-123 bp ladder

GenBank reference sequence. The domestic sheep matched the Goral (*Naemorhedus caudatus*) with an e-value of e^{-106}, and there were 21 differences between the experimental and the known sequences. Finally, the top BLAST match for the domestic dog had an e-value of 7 e^{-90} and corresponded to the Eastern African black-backed jackal cyt b gene with 29 differences from its respective GenBank sequence. Although the correct species was not identified for some specimens, in no instance did the BLAST result fail to associate the tested sequence with the correct family.

Phylogenetic Analysis

Three distance-based phylogenetic trees were generated , and the results were directly compared to the BLAST results. The first tree (Fig 10A) was generated using the ninety-four vertebrate cyt b sequences compiled from GenBank to demonstrate that all species' sequences were placed within proper classification groups (see Materials and Methods Database Development section). The results from the pair-wise comparison established that all clades were formed as expected based on class, order, and family classifications with the exception of the order rodentia (refer to Table 2 in Materials and Methods). The tree was subjected to bootstrap analysis with resulting bootstrap values ranging between 51 and 100.

The second phylogenetic tree compared the eleven single source experimental sequences with all ninety-five GenBank database sequences (Fig. 10B). Results demonstrated that a 100 percent confidence level was achieved for all experimental sequences, with the exception of marmoset (64%), and domestic horse (66%). The marmoset (exact species unknown) aligned with the common marmoset species sequence, and the domestic horse sequence was positioned next to the Equidae family.

These two sequences also displayed the lowest confidence during the BLAST searches. Regardless, both were placed with the correct family using both the phylogenetic and BLAST methods.

To generate the final tree, the 11 GenBank reference sequences corresponding to the species tested for the validation were removed so that the generated tree only compared the eleven experimental sequences to the remaining 83 GenBank reference sequences (Fig. 10C). For example, the cow database sequence was not included so that only the experimental cow sequence would be included. The first and second trees were compared (Fig. 10A) to determine whether the placement of the experimental sequences differed from the placement of the corresponding database sequences. Clade formations were the same with slight differences in arrangement, including combining two branches in a clade into one branch or differences in species order from top to bottom in the tree. For example in Figure 10B, the black howler monkey (*Alouatta caraya)*, Panamanian red spider monkey (*Ateles geoffroyi panamensis*), and the common marmoset formed a single group which then directly connected with the black-headed uakari (*Cacajao melanocephalus*) species. In Figure 10C, the black howler monkey and the marmoset sequences formed a branch pair, which then connected to the Panamanian red spider monkey sequence, and this group of three branched with the black-headed uakari. Bootstrap values for the eleven species tested were all above 50 with no species differing by >7% from the corresponding GenBank species sequences in the second tree. The lowest branch confidence was for placement of the domestic horse, which also displayed the highest BLAST e-value.

Fig. 9. Sequences of Contaminated Products. Three bases that are representative of the nature of each mixture are displayed for each species sequence. Letters above the peaks denote the base called by the computer (A, G, T, or C). The title of each figure indicates the major sequence contributor. Note the ratio of the smaller peaks to the larger peaks. **(A)** House mouse. **(B)** Domestic sheep. **(C)** Domestic dog.

Table 4. BLAST and Phylogenetic Analysis Summary. The number of experimental to reference sequence differences, the sequence length, the information pertaining to the top BLAST match for each species, including the common name, e-value, accession number, and gene; and the bootstrap value from the generated tree, are shown.

Name	Lgth.	TR	Top BLAST species	e-value	Accession #	Gene	Bootstrap
Alligator	307	0	Alligator	e^{-171}	AF318563.1	cyt b	100
Brown kiwi	305	0	Brown kiwi	e^{-165}	AY016010.1	mt genome	100
Chicken	307	0	Chicken	e^{-171}	AJ401080.1	cyt. b	100
Domestic cat	307	2	Domestic cat/Wild cat	e^{-171}	AY170102.1/AJ441328.1	cyt.b/cyt. B	100
Domestic cow	307	3	Domestic cow	e^{-171}	AF493542.1	cyt. b	100
Domestic pig	307	1	Domestic pig	e^{-171}	AF486866	mt genome	100
European rabbit	307	2	European rabbit	e^{-169}	AJ001588.1	mt genome	100
Gorilla	307	2	Gorilla	e^{-167}	D38114.1	mt genome	100
Human	307	1	Human	e^{-171}	AF346986.1	mt genome	100
Marmoset	298	52	Cotton-topped Tamarin	e^{-164}	AF001930.1	cyt. b	64
Domestic horse	307	44	Grevyi's zebra	$2e^{-71}$	X56282.1	cyt. b	66
House mouse	307	1	House mouse	e^{-171}	AY339599.2	mt genome	N/A
Domestic sheep	307	21	Gobi argali	e^{-106}	U17861	cyt. b	N/A
Domestic dog	306	29	Eastern African black-backed jackal	$7e^{-90}$	AF028143.1	cyt. b	N/A

Fig. 10. PAUP Generated Phylogenetic Trees. The outgroup (or root) for all trees was the rainbow trout *(Onchorynchus mykiss)* GenBank sequence. For class, order, and family classification, refer to Table 2. The values displayed on each branch of the trees are the bootstrap values. Trees generated from: **(A)** GenBank database sequences. **(B)** Experimental (tested during the current validation) and GenBank database sequences. The suffix pcr designates species tested at AFDIL during the current validation. For example, the GenBank alligator sequence and the experimental alligator sequence are designated Alligator mississippiensis and A. mississippiensis_pcr, respectively. **(C)** Experimental and database sequences.

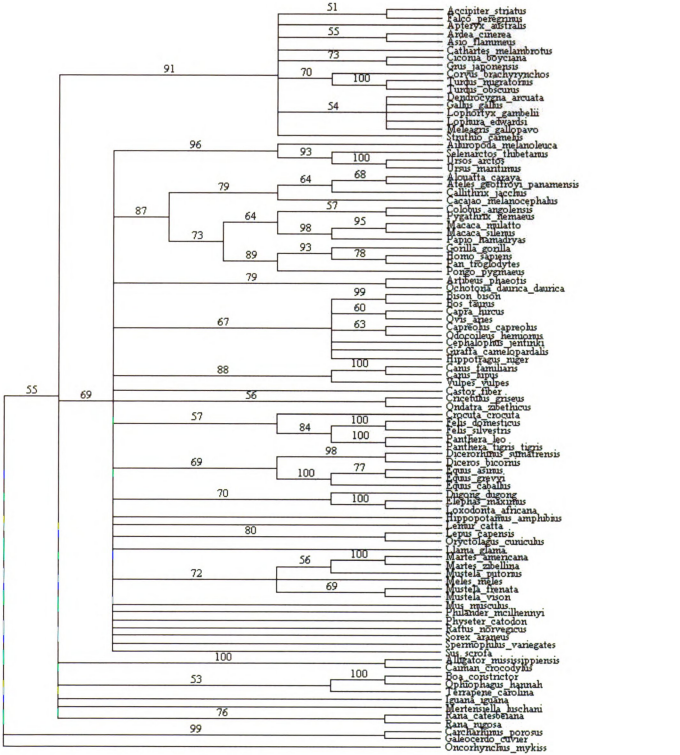

Accipiter_striatus
Falco_peregrinus
Apteryx_australis
Ardea_cinerea
Asio_flammeus
Cathartes_melambrotus
Ciconia_boyciana
Gnu_japonensis
Corvus_brachyrynchos
Turdus_migratorius
Turdus_obscurus
Dendrocygna_arcuata
Gallus_gallus
Lophortyx_gambelii
Lophura_edwardsi
Meleagris_gallopavo
Struthio_camelus
Ailuropoda_melanoleuca
Selenarctos_thibetanus
Ursos_arctos
Ursus_maritimus
Alcaelfa_cervus
Ateles_geoffroyi_panamensis
Callithrix_jaccinus
Cacajao_melanocephalus
Colobus_angolensis
Pygathrix_nemaeus
Macaca_mulatto
Macaca_silenus
Papio_hamadryas
Gorilla_gorilla
Homo_sapiens
Pan_troglodytes
Pongo_pygmaeus
Artibeus_phaeotis
Ochotona_daurica_daurica
Bison_bison
Bos_taurus
Capra_hircus
Ovis_aries
Capreolus_capreolus
Odocoileus_hemionus
Cephalophus_jentinki
Giraffa_camelopardalis
Hippotragus_niger
Canis_familiaris
Canis_lupus
Vulpes_vulpes
Castor_fiber
Cynocephalus_griseus
Ondatra_zibethicus
Capcula_crocuta
Felis_domesticus
Felis_silvestris
Panthera_leo
Panthera_tigris_tigris
Dicerorhinus_sumatrensis
Diceros_bicornis
Equus_asinus
Equus_grevyi
Equus_caballus
Dugong_dugong
Elephas_maximus
Loxodonta_africana
Hippopotamus_amphibius
Lepus_capensis
Oryctolagus_cuniculus
Lama_glama
Martes_americana
Martes_zibellina
Mustela_putorius
Mustela_nigripes
Mustela_frenata
Mustela_vison
Mus_musculus
Ephilander_mcilhennyi
Physeter_catodon
Rattus_norvegicus
Sorex_araneus
Spermophilus_variegates
Sus_scrofa
Alligator_mississippiensis
Caiman_crocodylus
Boa_constrictor
Ophiophagus_hannah
Terrapene_carolina
Iguana_iguana
Mertensiella_luschani
Rana_catesbeiana
Rana_pugosa
Gavialis_gangeticus
Caiman_crocodylus
Alligator_mississippiensis
Oncorhynchus_mykiss

Fig. 10A.

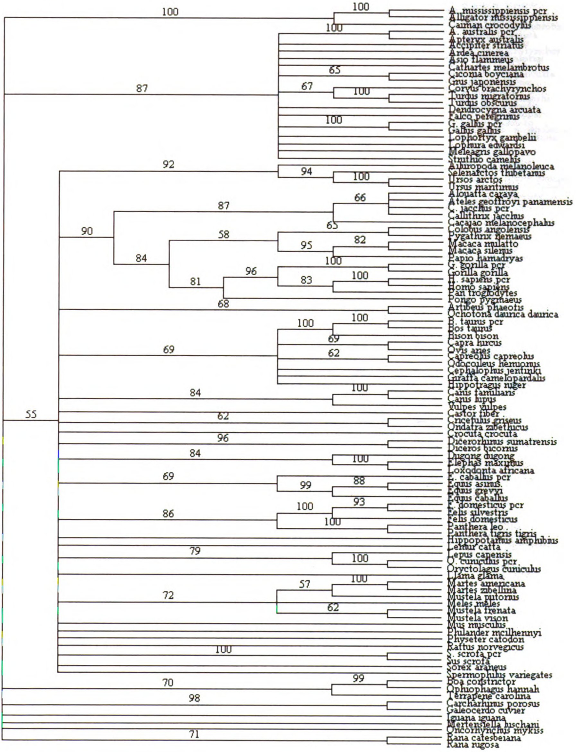

A. mississippiensis pcr
Alligator mississippiensis
Caiman crocodylus
A. australis pcr
Apteryx australis
Accipiter striatus
Ardea cinerea
Asio flammeus
Cathartes melambrotus
Ciconia boyciana
Grus japonensis
Corvus brachyrynchos
Turdus migratorius
Turdus obscurus
Dendrocygna arcuata
Falco peregrinus
G. gallus pcr
Gallus gallus
Lophortyx gambelii
Lophura edwardsi
Meleagris gallopavo
Struthio camelus
Ailuropoda melanoleuca
Selenarctos thibetanus
Ursus arctos
Ursus maritimus
Alouatta caraya
Ateles geoffroyi panamensis
C. jacchus pcr
Callithrix jacchus
Cacajao melanocephalus
Colobus angolensis
Pygathrix nemaeus
Macaca mulatto
Macaca silenus
Papio hamadryas
G. gorilla pcr
Gorilla gorilla
H. sapiens pcr
Homo sapiens
Pan troglodytes
Pongo pygmaeus
Artibeus phaeotis
Ochotona daurica daurica
B. taurus pcr
Bos taurus
Bison bison
Capra hircus
Ovis aries
Capreolus capreolus
Odocoileus hemionus
Cephalophus jentinki
Giraffa camelopardalis
Hippotragus niger
Canis familiaris
Canis lupus
Vulpes vulpes
Castor fiber
Cricetulus griseus
Ondatra zibethicus
Crocuta crocuta
Dicerorhinus sumatrensis
Diceros bicornis
Dugong dugong
Elephas maximus
Loxodonta africana
E. caballus pcr
Equus asinus
Equus grevyi
Equus caballus
F. domesticus pcr
Felis silvestris
Felis domesticus
Panthera leo
Panthera tigris tigris
Hippopotamus amphibius
Lemur catta
Lepus capensis
O. cuniculus pcr
Oryctolagus cuniculus
Llama glama
Martes americana
Martes zibellina
Mustela putorius
Meles meles
Mustela frenata
Mustela vison
Mus musculus
Philander mcilhennyi
Physeter catodon
Rattus norvegicus
S. scrofa pcr
Sus scrofa
Sorex araneus
Spermophilus variegates
Boa constrictor
Ophiophagus hannah
Terrapene carolina
Carcharhinus porosus
Galeocerdo cuvier
Iguana iguana
Mertensiella luschani
Oncorhynchus mykiss
Rana catesbeiana
Rana rugosa

Fig. 10B.

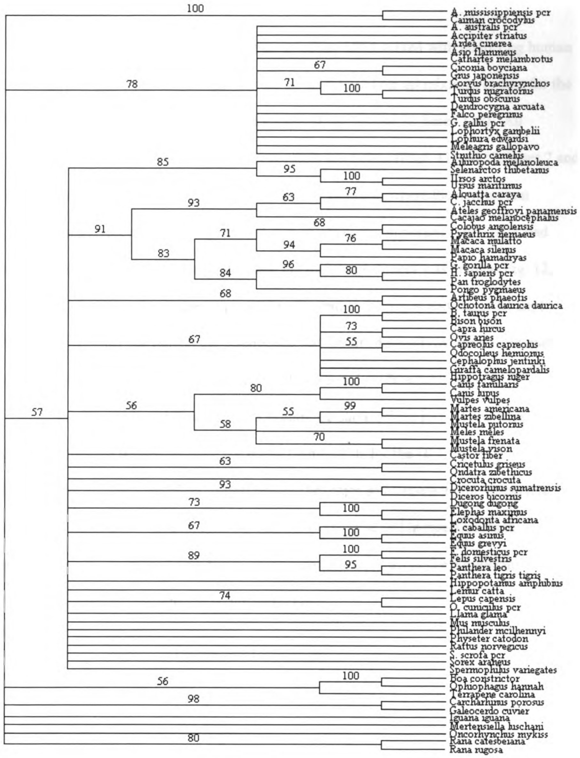

A. mississippiensis pcr
Caiman crocodylus
A. australis pcr
Accipiter striatus
Ardea cinerea
Asio flammeus
Cathartes melambrotus
Ciconia boyciana
Grus japonensis
Corvus brachyrynchos
Turdus migratorius
Turdus obscurus
Dendrocygna arcuata
Falco peregrinus
G. gallus pcr
Lophortyx gambelii
Lophura edwardsi
Meleagris gallopavo
Struthio camelus
Ailuropoda melanoleuca
Selenarctos thibetanus
Ursos arctos
Ursus maritimus
Alouatta caraya
C. jacchus pcr
Ateles geoffroyi panamensis
Cacajao melanocephalus
Colobus angolensis
Pygathrix nemaeus
Macaca mulatto
Macaca silenus
Papio hamadryas
G. gorilla pcr
H. sapiens pcr
Pan troglodytes
Pongo pygmaeus
Artibeus phaeotis
Ochotona daurica daurica
B. taurus pcr
Bison bison
Capra hircus
Ovis aries
Capreolus capreolus
Odocoileus hemionus
Cephalophus jentinki
Giraffa camelopardalis
Hippotragus niger
Canis familiaris
Canis lupus
Vulpes vulpes
Martes americana
Martes zibellina
Mustela putorius
Meles meles
Mustela frenata
Mustela vison
Castor fiber
Cricetulus griseus
Ondatra zibethicus
Crocuta crocuta
Dicerorhinus sumatrensis
Diceros bicornis
Dugong dugong
Elephas maximus
Loxodonta africana
E. caballus pcr
Equus asinus
Equus grevyi
F. domesticus pcr
Felis silvestris
Panthera leo
Panthera tigris tigris
Hippopotamus amphibius
Lemur catta
Lepus capensis
O. cuniculus pcr
Llama glama
Mus musculus
Philander mcilhennyi
Physeter catodon
Rattus norvegicus
S. scrofa pcr
Sorex araneus
Spermophilus variegates
Boa constrictor
Ophiophagus hannah
Terrapene carolina
Carcharhinus porosus
Galeocerdo cuvier
Iguana iguana
Mertensiella luschani
Oncorhynchus mykiss
Rana catesbeiana
Rana rugosa

Fig. 10C.

Mixture Studies

Amplification of up to 100 pg of invertebrate DNA mixed with up to 1 pg human DNA showed that amplification product could be detected for all mixtures except for the 100 pg invertebrate:0 pg human DNA reactions (Figs. 11 and 12, lanes 3–9). All amplification controls and the specificity controls were clear (Figs. 11 and 12, lanes 2 and 10 and 2 and 11–12, respectively). The vertebrate mixtures where the non-human component was included at a higher ratio than the human component (10:1–3:2) had bands of greater intensity than those where the major constituent was human (Fig. 12, lanes 3–9). All reactions were sequenced to determine the major and minor (if any) components. Only the human cyt b sequence was obtained for all invertebrate:human mixtures, except for the 100:0 mixture, which produced no detectable sequence.

Results of the vertebrate mixture studies are summarized in Table 5. Most non-human vertebrate sequences were the major components for the 10:1 to 2:3 (non-human:human) ratios based on comparison of the major sequence with the respective reference sequences. For example, the American alligator and human GenBank sequences were compared with the major component sequence of the American alligator:human mixtures to see which matched.

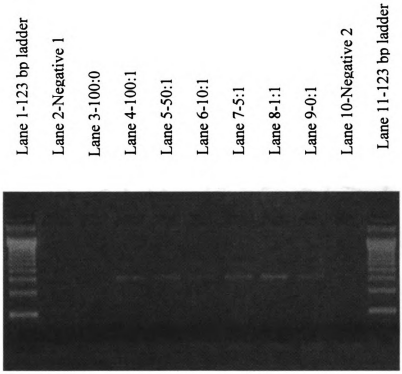


Fig. 11. Example of an Invertebrate:Vertebrate Mixture Amplification. The image is of the dilution reactions from the yeast:human mixtures. Lane numbers and dilution values are designated at the top of the figure.

Fig. 12. Example of a Non-human vertebrate:human Mixture Product Gel. Agarose gel of American alligator:human amplification. Lane numbers and dilution values are designated at the top of the figure.

Table 5. Mixture Results Summary Table. The primary contributor for each mixture is indicated in the table. Any mixtures in which a primary and secondary profile could not be distinguished were labeled as inconclusive.

	10:0	10:1	9:1	3:2	1:1	2:3	1:9	1:10	0:10
Alligator:Human	A	A	A	I	A	A	H	H	H
Chicken:Human	C	C	C	C	C	C	C	H	C
Gorilla:Human	G	G	G	I	H	H	H	H	H

*A = alligator as primary, C = chicken, G = gorilla, H = human, I = inconclusive

DISCUSSION

It has been the United States Military policy to have a full accounting of all service members that are missing or and killed in action. Towards this end, the CILHI is charged with the responsibility of recovering these remains from Korea, Vietnam or any World War II site and identifying them so they can be returned to their families (reviewed by Holland and Parsons 1999). The nature of the incident, the environment, and the time since death all influence the state of the skeletal remains. In instances where the individual died in a high impact crash (e.g. airplane) or explosion, intact pieces of bone as well as highly fragmented bone that are not anthropologically identifiable as human may be submitted to the AFDIL for mtDNA testing. This leads to problems with determining whether amplification failures result from inhibition or to the extracts being from non-human samples.

Amplification Optimization and Sequencing

To address if the specimen was human or non-human, two visiting George Washington University graduate students conducted a preliminary cyt b study at AFDIL in 2000. During the course of this preliminary investigation, Parson et al.'s (2000) mitochondrial cyt b primers were used to amplify and identify DNA extracts from several vertebrate species. The students used 25 – 35 cycles to amplify 5 pg or more of genomic vertebrate DNA.

Despite the success of the initial study, AFDIL requires a limit of detection of 1 pg to validate any new primers for use in mtDNA testing. The current validation demonstrated that a 1 pg sensitivity was achieved for the 9 quantified vertebrate species

tested using the Parson et al. (2000) parameters (cyto 2) with 42 amplification cycles instead of the 30 or 35 cycles used by the authors with the addition of a 7 minute extension step. Although maintaining lower amplification cycles is usually preferred to prevent non-specific amplification, AFDIL scientists have demonstrated that amplification of low copy number or degraded DNA extracts could be achieved when smaller regions were amplified with 38 to 42 cycles (Fisher et al. 1993)

The primers were confirmed as vertebrate specific when 100 pg of seven species of invertebrate DNA produced no detectable amplification product, and only the human DNA was detected for all of the invertebrate:human mixtures including when invertebrate DNA was present at a 100 times higher concentration. Interestingly, though product gel bands were present for all vertebrate species at 42 cycles, differences in band intensity were noted among the 1 pg products (Fig. 7C). The 1 pg gorilla extract produced a brighter band than all the other species including human, while the alligator extract gave the weakest band intensity. Mixture results also revealed differing PCR efficiencies instead of equally intense product gel bands for all mixture reactions and electropherogram peaks of equal heights for both vertebrate species. Brighter bands on agarose gels were present for most of the higher non-human:human vertebrate mixture ratios (e.g. – 10:1, 9:1, 3:2). The fainter bands for the lower ratios (1:9, 1:10) were an indication that human DNA was not amplified as efficiently as the non-human species' DNA. Branicki et al. (2003) also found differences in PCR efficiency when observing product gel results for amplification of cow and pig DNA dilution series. These authors demonstrated that pig DNA amplified more efficiently than cow DNA, and the differences were assumed to be related to the number of DNA sequence differences

present in the primer binding sites for each species (Branicki et al. 2003). However, no sequence data were provided to confirm this conclusion.

In an attempt to explain the differing amplification efficiencies, the current validation determined the number of primer binding site differences compared to the published GenBank reference sequence for each species tested. This assessment revealed that as little as one and at the most six sequence differences existed. However, no direct correlation between the number of primer binding site variations and the intensity of the agarose gel bands was indicated. For example, the gorilla, with the greatest number of primer binding site differences, might be expected to have the least intense band when compared to the other tested species, however it was the brightest. Furthermore, the positioning of the sequence differences did not seem to influence the amplification efficiency, which tended to be interspersed throughout the forward and reverse sequences. Concentration of the sequence differences at the 3' end(s) of the forward and/or reverse primers could potentially reduce the primer binding efficiency, this was not present for the set of tested species sequences. Even the species that had one or two differences at the 3' end (e.g. – alligator, cow, gorilla) showed similar or greater agarose band intensities than species that had no differences in that region (e.g. - human, chicken, pig). For example, the alligator had the same number of primer differences as cat and cow with one of the differences at the second to last base from the 3' end of the reverse primer (5). Yet, the cow, which had a greater number of sequence differences at the 3' ends of both primers, generated a far more intense band than the alligator.

An alternate explanation for the variation in intensity is that the original DNA concentrations were incorrect. This seems unlikely because the manufacturer provided

the concentrations for the BIOS laboratory specimens, and all other specimens (except for the kiwi) had been quantified by UV spectrophotometry, however slight variation could certainly exist. Other factors, including the amount of mtDNA contained within a sample (as opposed to total DNA, which is measured using spectrophotometry), or DNA secondary structures such as hairpin formation in the template DNA that interfere with PCR, could also affect results.

Intraspecies differences were often found between the experimental amplified cyt b sequences and the corresponding published GenBank sequences (Table 4) during this validation. Previous cyt b research showed that intraspecies variation is encountered during comparisons of sequences from multiple representatives of the same species, attributable to the normal mutation of mtDNA (Cronin et al. 2001, Hsieh et al. 2001 and 2003). Any intraspecies differences do not seem to interfere with species identification though. For example, Hsieh et al. (2001) found that the percentage of intraspecies sequence differences for 19 vertebrate species tested ranged from 0.25 to 2.74%, far lower than the 5.97 to 34.83% percentage of interspecies differences.

Sequence results also revealed that 3 of the 14 species DNA samples (house mouse, domestic sheep, and domestic dog) were contaminated with a different species The contaminating species were not identified though the major contributing sequence was separated from the minor by making a visual determination of the major base at each position in the sequence and manually adjusting the sequence according to that determination. For example, if both an A and a C were at one position but the C had a lower peak height, the major peak at that position was called an A. If both the A and C appeared to have equal heights, the major base could not be determined and the peak was

called an N. Contamination of the house mouse specimen most likely occurred during previous use of the specimen since none of the other extracts, amplification control, or specificity control PCR reactions set up at the same time were contaminated. The contaminant peaks were so low (<1% of the major contributor peaks) that elevated baseline could not be ruled out, but to be conservative the peaks were considered to be those of a low level contaminant. This meant that the major sequence was isolated, compared to the mouse reference sequence, and imported into BLAST for a species determination, but the sequence was not included in the phylogenetic tree generation. Unlike the mouse, the domestic sheep and domestic dog specimens obtained from the University of Delaware were highly contaminated (>50% of the major peaks). Contamination of these specimens most likely occurred at the time of collection of the blood samples or at the time of packaging of the FTA® blood cards and not during the extraction procedure. This was further supported by the lack of contamination of two of the other extracts from this group (domestic cat and domestic horse) as well as the reagent blank, which were all extracted at the same time. Extraction of a new sample from the FTA® blood cards of these 4 specimens confirmed that contamination had not occurred during the extraction procedure and that the BLAST results for all four species were reproducible. Species identification of the contaminated samples was still attempted because the possibility for contamination during casework analysis does exist though the occurrence is extremely rare.

BLAST Identification

The low level of contamination did not influence BLAST based species identification in the house mouse; the top BLAST match was the house mouse cyt b gene

with a value of e^{-171}. In contrast, because of the large number of N's interspersed throughout the sequences, successful BLAST identification was not achieved for the contaminated domestic sheep and domestic dog sequences though the top BLAST matches for each were in the proper family. The two sequences may have been identified correctly had they been single sources, though Branicki et al. (2003) reported no instances of contamination and found that a BLAST search was unable to distinguish between amplicons of mouflon sheep (*Ovis musimon*) and domestic sheep or between wolf (*Canis lupus*) and domestic dog.

One should keep in mind when assessing the BLAST results that all of the species tested were known to be in the GenBank database. The development of an exhaustive reference database comprising the foreign species (e.g. from Vietnam and Korea) that could potentially be encountered would require substantial time and resources. Fortunately, the need for such a database is superceded by the large number of vertebrates that are currently encountered in GenBank. For example, Branicki et al. (2003) found that cyt b sequence data for three of the 34 species they tested could not be found in the database, but the species were able to be matched with closely related species that were in GenBank. In addition, Parson et al. (2000) found that the only types of vertebrate cyt b sequences that could not be found in GenBank at the time of their study were avian. In compiling the 94 database species' sequences for the current project, avian and amphibian cyt b species sequences were less common in the GenBank database. This presents an obstacle only when exact species identification is necessary. For this validation, exact species identification is advantageous but not necessary since the desired result is a non-human versus human designation.

Other BLAST discrepancies encountered during the course of this validation were associated with identification of the domestic cat, marmoset, and the domestic horse. The domestic cat cyt b sequence matched equally well with the wild cat (*Felis silvestris*) and domestic cat GenBank cyt b sequences. Branicki et al. (2003) reported the same result, noting that the two species are indistinguishable based on cyt b sequence data alone. The marmoset and domestic horse, neither of which was correctly identified, also presented interesting results. The exact marmoset species used in this study was unknown, but the sequence was a 99% match to a tamarin sequence instead of any of the marmoset species sequences. In considering explanations for the incorrect match, it was noted that only partial cyt b sequences were available in GenBank for all members of the Callithrichidae family except the common marmoset. For example, only 255 bases of Snethlage's marmoset (*Callithrix emiliae*) were available for comparison only 255 bases for the cyt b gene of which 214 (81%) overlapped with the entered marmoset sequence. In addition, not all marmoset species are represented in the GenBank database. As a result, misidentification may have been based on the absence of the correct species from the GenBank database.

The horse exhibited a large >50 sequence differences in the comparison to the GenBank horse sequence, and the top BLAST match was the zebra cyt b sequence. A reasonable explanation for these results has yet to be determined. The sequence was clearly from a single source and originated from a domestic horse based on the labeling of the FTA® specimen received from the University of Delaware. The possibility that the sequence was that of a nuclear pseudogene (insertion of the cytochrome b gene sequence into the nuclear genome) could be considered according to the characteristics

outlined by Irwin et al (1991). Irwin et al. (1991) listed the following characteristics: the presence of two peaks at many sequence positions with no contaminant present, a large number of base substitutions compared to the expected number of substitutions at each codon position for mtDNA, and a lower than expected ratio of transitions at third codon positions to transitions at first codon positions (Mundy et al. 2000). The specimens were not evaluated for presence of those characteristics, but two other properties indicative of pseudogenes, indeterminate sequence length (Irwin et al. 1991) and presence of stop codons in all reading frames (Johns and Avise 1998, Mundy et al. 2000), were not observed.

In summary, the effectiveness of the technique was demonstrated when nine of the 11 (81%) non-contaminated samples were matched with the mitochondrial genome or cyt b gene sequence of the corresponding GenBank species. When contaminated sequences were included 10 of 14 (71%) of the species were correctly identified with 100% of the sequences associated with the correct family.

Phylogenetic Tree Comparisons

Three phylogenetic trees were generated using PAUP as described in Methods and Materials. The first was used to evaluate the accuracy of clade formation using the known compilation of 94 species sequences compared to Table 2. All clades were formed as expected with the exception of the rodents; members of the order Rodentia did not form a single clade. The hamster and muskrat branched out from the same node to form a cluster, which was positioned adjacent to the muskrat branch, while the remaining rodent species occupied their own independent branches further down the tree. Investigation into the evolutionary relationships among rodent species revealed an

ongoing debate concerning the monophyly (or lack thereof) of the order. Authors including Huchon et al. (2002) and Sullivan and Swofford (1997) have asserted that the monophyly of rodents has yet to be disproved. On the other hand, Graur et al. (1992) and Li et al. (1992) discussed the paraphyly of Rodentia, saying that the rodents branch off into separate groups including guinea-pig-like rodents (caviomoprhs) and rat-like rodents (myomorphs). This was supported by the observations of Reyes et al. (2000), who indicated that rodents are either polyphylectic or paraphylectic based on placement of several rodent species within a mammalian phylogenetic tree. Likewise, the work of Honeycutt et al. (1995) supported rodent polyphyly when the cyt b gene sequences of 35 mammalian species were aligned.

The results of the validation described here also support the theory of evolutionary separation of the rodent order. One should understand, however, that the tree generation criteria were extremely conservative; certain assumptions, such as equal transversion and transition rates, no species-based bias towards transversions or transitions, and equal frequencies for each base, were made. Adjusting these with more specific values would result in a slightly different and possibly more accurate evolutionary tree (Honeycutt et al. 1995, Huelsenbeck 1995, Irwin et al. 1991, McClellan and McCracken 2001), but such adjustments were beyond the scope of this project.

The second tree, generated under the same conditions as the first, was used to determine whether each tested species sequence grouped with its respective GenBank reference sequence. All experimental species aligned with the proper sequence with bootstrap values of 100 except for the domestic cat (93%), the marmoset (66%), and the domestic horse (69%). The experimental domestic cat and the GenBank wild cat

sequences branched together, once again supporting the inability to distinguish the two species by cyt b sequence comparison though the experimental domestic cat/GenBank wild cat cluster did branch with the GenBank domestic cat sequence with a bootstrap value of 100%. Though the latter two species, marmoset and domestic horse, both grouped with the correct species in the phylogenetic tree, the associated bootstrap values are in keeping with the BLAST search results as they also had the lowest confidence (based on e-values) for the top BLAST matches. The alignments of the marmoset with the common marmoset sequence and the domestic horse with the domestic horse sequence should be evaluated with caution. One should keep in mind, for example, that as with BLAST, only a limited number (one in this case) of Callithricidae family sequences is available in the 94 database sequences.

Finally, a third tree was generated to compare the alignment pattern of the experimental sequences with the corresponding GenBank sequences in the second tree (Fig. 10C versus Fig. 10B). The generation of the third tree was necessary since sequence differences within a species could indirectly affect the arrangement and confidence values for other branches of the tree. This is a consideration since the horse and marmoset sequences differed from their respective GenBank sequences by over 50 base pairs. The lower level of confidence for the placement of these two species without the presence of the reference sequences could affect the values for placements further out in the tree. For example, because the horse bootstrap confidence was lower, the positioning of the next branch out may have been lower and so on. Upon assessment of the third generated tree, the branching arrangement was the same as for the second tree, and most placements were either identical or 2–3% lower in confidence than for the

second tree, with the greatest difference being a 7% (higher) difference for the placement of the domestic cat sequence. The sequence aligned with the wild cat sequence with a 93% bootstrap value in Fig. 10B, but the bootstrap value was 100% (the same as in Fig. 10A) in Fig. 10 C because the sequences are indistinguishable and only two of the three (GenBank domestic cat, experimental domestic cat, and GenBank wild cat) sequences were being compare in the third tree.

The three phylogenetic trees depicted in the results were compiled without the contaminated sequences to prevent skewing of results caused by the large number of uncalled (N) bases. A separate alignment was evaluated in PAUP using all sequences, including the contaminated house mouse, domestic sheep, and domestic dog sequences (data not shown). All alignments were the same, except the bootstrap values for gorilla and human were 99 and 98 respectively instead of the original 100 percent. The dog sequence aligned with the experimental brown kiwi sequence with a bootstrap value of 86. The large number of ambiguous bases in the dog sequence are most likely the reason for this misalignment. The sheep sequence aligned with the GenBank domestic sheep sequence with a bootstrap value of 57. This correct alignment, despite the 21 sequence differences, can likely be attributed to the single ovid species available for comparison in the database generated for this study. As with BLAST, the high (100%) confidence of the experimental and control house mouse sequence alignment was likely a result of the low level of the contaminating DNA and the absence of ambiguous bases in the major sequence.

BLAST Versus Phylogenetic Tree Comparison

A comparison of BLAST searching and phylogenetic alignment was undertaken

to determine the more efficient and/or accurate method of species ID. The BLAST comparison and phylogenetic alignment could both be used for species identification , but the BLAST program was chosen for the final validated procedure for two major reasons. First, the number of known sequences being compared through BLAST is far more than the number compiled for the reference database for this project (>35,000 versus 94). As discussed above, the greater number of sequences for comparison adds more weight to the confidence values for matches. Second, BLAST comparison is more time efficient for laboratories because it eliminates the need to develop an internal reference database for phylogenetic alignment.

Mixture Analysis

Non-human vertebrate:human mixtures were evaluated for separation of the major and minor components. Minor component RFU's that were less than 50% of the major component allowed the separation of sequences for all mixtures except the 3:2 alligator:human and 3:2 gorilla:human mixtures that were indistinguishable. The chicken DNA completely dominated the mixture reaction as it was the primary sequence for all but the 1:9 mixture. These results are in keeping with those of Branicki et al. (2003), who found that the cyt b DNA of some species is more readily detected when sequencing the amplification products of mixture reactions than the DNA of other species. For example, when analyzing a pig to human mixture series, the group found that a clear human signal was detected for six of the seven ratios, and the pig DNA was only detected by itself at the 1:100 ratio. During analysis of dog:pig mixture, a "pure" dog signal was never observed at any dilution, and the dog signal was only evident in two of the dilutions (100:1 and 50:1). Though the authors attributed the differences in efficiency to

primer binding site variations, they failed to address other possibilities, including starting amounts of mtDNA existing in their samples.

Based on Branicki et al.'s (2003) results and the results obtained in the present study, caution should be taken when evaluating mixture sequences. The major species in a contaminated specimen may actually appear to be the minor component if it amplifies less efficiently, but there is no way of determining whether this occurs. Therefore, whenever possible, the apparent minor component sequence should be determined in addition to the apparent major component sequence. It should also be noted that in this study a total of 1 pg of DNA (including both the major and minor components) was used for all mixture reactions. The low amount of DNA may have increased the potential for amplification of one species over the other because of the limited amount of template available in the reaction for each species. The importance of such an occurrence for AFDIL is likely limited however because the outer cortical layer of the bone is removed before the DNA is extracted (Armed Forces DNA Identification Laboratory DNA Extraction Manual, Version 2.0: "Organic Extraction of DNA from Dried Skeletal Remains"). The possibility of competitive amplification has to be considered, regardless, to account for the rare instance of contamination with an analyst's DNA.

The Validated Procedure

The procedure resulting from this research is advantageous to AFDIL for several reasons. First and foremost is the ease of implementation of the methods, which would require only minimal training for mtDNA analysts. Second, the procedure was developed around the current amplification conditions of the mini primer sets, using similar cycle numbers and achieving the same sensitivity. This is advantageous because amplification

failure would still provide valuable information since lack of amplified product using cyt b primers would be an indication of inhibition. Analysts could then proceed with efforts to address inhibition, such as diluting the template or adding more BSA. Third, the BLAST database is a readily available source of references for the vertebrate cyt b gene making identification as simple and quick as inserting a sequence into BLAST and awaiting the results (~ 1 min. or less).

Future Considerations

A final important aspect of any forensic validation is to determine if the developed procedure is applicable to case quality specimens. All non-contaminated extracts evaluated during the course of this study came from relatively rich and pristine sources of DNA, which may behave differently than small, possibly degraded skeletal specimens. Therefore, casework certified mtDNA analysts are currently re-extracting skeletal remains including some that were previously submitted to AFDIL by CILHI but failed to yield amplified product with human specific primer sets and mini-primer sets. The cyt b primers will be used to determine whether the unsuccessful amplification was a result of the bone being non-human or a result of severe degradation (Timothy McMahon, Ph.D., personal communication).

One extension of this project that may be beneficial to AFDIL is to evaluate the potential use of a multiplex comprised of human-specific D-loop primers and the separate vertebrate-specific cytochrome b primers so the human/non-human differentiation may be made solely via evaluation of agarose gel results, bypassing the need for DNA sequencing. Human samples would have two bands in this instance, and non-human vertebrates would only have the band corresponding to the cytochrome b amplicon.

Bellis et al (2003) used such a procedure to distinguish goat, cow, sheep, tiger, horse, cat, chicken, dog, and pig from human. However, in their study, the dog also produced two bands (cause unknown), though the positioning differed enough from the human bands to distinguish the two. Using this procedure could add greater efficiency to species differentiation if it is effective for ancient skeletal remains since further sequence and BLAST analysis would only be necessary when and if a specific species needed to be determined.

In conclusion, small and/or degraded bone fragments received at AFDIL are first amplified with primer set 2, which amplifies nucleotides 16190–16410 and is the most sensitive of the primer sets. Amplification failure with this primer set is followed by PCR using the most sensitive of the mini primer sets. If amplification is still ineffective, troubleshooting measures such as amplification with increased Taq, diluted template, or increased template is attempted. The validated procedure would potentially be used as the first step in troubleshooting to prevent wasting time on attempting to amplify non-human bones with human specific primers and seeing inhibition. The final validated protocol to be implemented is as follows: 1) 42 cycle amplification using the cyt b primers, 2) agarose gel electrophoresis to verify amplification, 3) purification and sequencing with the cyt b primers, and 4) import of the sequence into BLAST for identification. Note that for the validated procedure, scientists will simply copy the consensus sequence to BLAST and determine the species based on the top match for the search; no comparison to a known sequence using an alignment program such as Sequencher will be necessary. Though the ability to identify species to the family level is more than sufficient for AFDIL and other human forensic DNA laboratories, wildlife

forensic scientists may require greater discrimination. Therefore, BLAST searches may not always be specific enough. In these cases, the development of a separate internal database would be beneficial. For example, if one is interested in species identification of twenty species of felids, it could be necessary to obtain reference sequences for the them and perform a phylogenetic tree analysis. The necessity for some forensic scientists to achieve more specific identification may also be addressed using the immunological or protein assays outlined in the introduction or by amplification and sequencing of the entire cytochrome b gene, depending on sample condition (Guglich et al. 1994, Hillis et al. 1994, Irwin et al. 1991).

The validation results presented here demonstrate that the cyt b primers were specific and usable for limited amounts of DNA. The procedure is currently being implemented for use at AFDIL for the previously mentioned test samples from CILHI and will be implemented for casework in the near future (Timothy McMahon, Ph.D., personal communication).

BIBLIOGRAPHY

Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P.; eds. 2002. *Molecular Biology of the Cell.* Garland Science: New York.

Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. 1990. Basic alignment search tool. *Journal of Molecular Biology* 215(3):403–410.

Anderson, S.; Bankier, A.T.; Barrell, B.G.; de Bruijn, M.H.L.; Coulson, A.R.; Drouin, J.; Eperon, I.C.; Nierlich, D.P.; Roe, B.A.; Sanger, F.; Schreier , P.H.; Smith, A.J.H.; Staden, R.; Young, I.G. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290: 457–464.

Andrasko, Jan and Rosen, Bjorn. 1994. Sensitive identification of hemoglobin in bloodstains from different species by high performance liquid chromatography with combined UV and fluorescence detection. *Journal of Forensic Sciences* 39(4): 1018–1025.

Arnberg, A.; von Bruggen, E.F.; Borst, P. 1971. the presence of DNA molecuoles with a displacement loop in standard mitochondrial DNA preparations. *Biochimica Biophysica Acta* 246(2):353–357.

Bartlett, S. and Davidson, W.S. 1992. FINS (Forensically Informative Nucleotide Sequencing): a procedure for identifying the animal origin of biological specimens. *Biotechniques* 12(3):408–411.

Bellis, C.; Ashton, K.J.; Freney, L.; Blair, B; and Griffiths, L.R. 2003. A molecular genetic approach for forensic animal species identification. *Forensic Science International* 134:99–108.

Blackett, R.S. and Keim, P. 1992. Big game species identification by deoxyribonucleic acid (DNA) probes. *Journal of Forensic Sciences* 37(2):590–596.

Bogenhagen, D. and Clayton, D.A. 1974. The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells: quantitative isolation of mitochondrial deoxyribonucleic acid. *The Journal of Biological Chemistry* 249(24):7991–7995.

Bose, S.; French, S.; Evans, F.J.; Joubert, F.; Balaban, R.S. 2003. Metabolic network control of oxidative phosphorylation: multiple roles of inorganic phosphate. *Journal of Biological Chemistry* 278(40):39155–39165.

Branicki, W.; Kupiec, T.; and Pawlowski, R. 2003. Validation of cytochrome b sequence analysis as a method of species identification. *Journal of Forensic Sciences* 48(1):83–87.

Brown, W.M.; George, M.; Wilson, A.C. 1979. Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences USA* 76(4):1967–1971.

Budowle, B.; Allard, M.W.; Wilson, M.R. 2002. Critique of interpretation of high levels of heteroplasmy in the human mitochondrial DNA hypervariable region I from hair. *Forensic Science International* 126:30–33.

Butler, J.M.; David, V.A.; O'Brien, S.J.; and Menotti-Raymond, M. 2002. The Meowplex: a new DNA test using tetranucleotide STR markers for the domestic cat. www.promega.com.

Cann, R.L.; Stoneking, M.; and Wilson, A.C. 1987. Mitochondrial DNA and human evolution. *Nature* 325:31–36.

Carracedo, A.; Bar, W.; Lincoln, P.; Mayr, W.; Morling, N.; Olaisen, B.; Schneider, P.; Budowle, B.; Brinkmann, B.; Gill, P.; Holland, M.; Tully, G.; and Wilson, M. 2000. DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. *Forensic Science International* 110:79–85.

Chang, D.D. and Clayton, D.A. 1985. Priming of human mitochondrial DNA replication occurs a the light strand promoter. *Proceedings of the National Academy of Sciences USA* 82(2):351–355.

Cronin, M.A.; Palmisciano, D.A.; Vyse, E.R.; and Cameron, D.G. 1991. Mitochondrial DNA in wildlife forensic science: species identification of tissues. *Wildlife Society Bulletin* 19:94–105.

D'Eustachio, P. 2002. High levels of mitochondrial DNA heteroplasmy in human hairs by Budowle et al. *Forensic Science International* 130:63–67.

Efron, B.; Halloran, E.; and Holmes, S. 1996. Bootstrap confidence levels for phylogenetic trees. *Proceedings of the National Academy of Sciences USA* 93:13429–13434.

Espinoza, E.O., Kirms, M.A., and Filipek, M.S. 1996. Identification and quantitation of source from hemoglobin of blood and blood mixtures by high performance liquid chromatography. *Journal of Forensic Sciences* 41(5):804–811.

Esposti, M.D.; De Vries, S.; Crimi, M.; Ghelli, A.; Patarnello, T.; and Meyer, A. 1993. Mitochondrial cytochrome b: evolution and structure of the protein. *Biochim Biophys Acta* 1143(3):243–71.

Fisher, D.L; Holland, M.M.; Mitchell, L.; Sledzik, P.S.; Wilcox, A.W.; Wadhams, M.; and Weedn, V.W. 1993. Extraction, evaluation , and amplification of DNA from

decalcified and undecalcified United States civil war bone. *Journal of Forensic Sciences* 38(1):60–68.

Folin, M. and Contiero, E. 1996. Electrophoretic analysis of non-human primates hair keratin. *Forensic Science International* 83(3):191–199.

Foran, D.R.; Crooks, K.R.; and Minta, S.C. 1997. Species identification from scat: an unambiguous genetic method. *Wildlife Society Bulletin* 25(4):835–839. a

Foran, D.R.; Minta, S.C.; and Heinemeyer, K.S. 1997. DNA-based analysis of hair to identify species and individuals for population research and monitoring. *Wildlife Society Bulletin* 25(4):840–847. b

Giles, R.E.; Blanc, H.; Cann, H.; and Wallace, D.C. 1980. Maternal inheritance of human mitochondrial DNA. *Proceedings of the National Academy of Sciences USA* 77(11):6715–6719.

Grzybowski, T. 2000. Extremely high levels of human mitochondrial DNA heteroplasmy in single hair roots. *Electrophoresis* 21:548–553.

Guglich, E.A.; Wilson, P.J.; and White, B.N. 1994. Forensic application of repetitive DNA markers to the species identification of animal tissues. *Journal of Forensic Sciences* 39(2):353–361.

Hall, B.G. 2001. *Phylogenetic trees made easy: a how-to manual for molecular biologists*. Sunderland, Mass., Sinauer Associates.

Hillis, D.M., Huelsenbeck, J.P.; Cunningham, C.W. 1994. Application and accuracy of molecular phylogenies. *Science* 264:671–676.

Holland, M.M.; Fisher, D.L.; Mitchell, L.G.; Rodriquez, W.C.; Canik, J.J.; Merril, C.R.; Weedn, V.W. 1993. Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam war. *Journal of Forensic Sciences* 38(3):542–553.

Holland, M.M. and Parsons, T.J. 1999. Mitochondrial DNA sequence analysis – validation and use for forensic casework. *Forensic Science Review* 11(1):21–50.

Honeycutt, R.L.; Nedbal, M.A.; Adkins, R.M.; and Janecek, L.L. 1995. Mammalian mitochondrial DNA evolution: a comparison of the cytochrome b and cytochrome c oxidase II genes. *Journal of Molecular Evolution* 40:260–272.

Hsieh, H.; Chiang, H.; Tsai, L.; Lai, S.; Huang, N.; Linacre, A.; and Lee, J.C. 2001. Cytochrome b gene for species identification of the conservation animals. *Forensic Science International* 122:7–18.

Hsieh, H.; Huang, L.; Tsai, L.; Kuo, Y.; Meng, H.; Linacre, A.; and Lee, J.C. 2003. Species identification of rhinoceros horns using the cytochrome b gene. *Forensic Science International* 136:1–11.

Huchon, D.; Madsen, Ole; sibbald, M.J.J.B; Ament, K.; Stanhope, M.J.; Catzeflis, F.; de Jong, W.W.; and Douzery, E.J.P. 2002. Rodent phylogeny and a timescale for the evolution of Glires: evidence from an extensive taxon sampling using three nuclear genes. *Molecular Biology and Evolution* 19(7):1053–1065.

Huelsenbeck, J.P. 1995. The robustness of two phylogenetic methods: four-taxon simulations reveal a slight superiority of maximum likelihood over neighbor joining. *Molecular Biology and Evolution.* 12(5):843–849.

Hutchison, C.A., III; Newbold, J.E.; Potter, S.S.; Edgell, M.H. 1974. Maternal inheritance of mammalian mitochondrial DNA. *Nature* 251(5475):536–538.

Ingman, M.; Kaessmann, H.; Paabo, S.; Gyllensten, U. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* 408(6813):708–713.

Irwin, D.M.; Kocher, T.D.; and Wilson, A.C. 1991. Evolution of the cytochrome b gene of mammals. *Journal of Molecular Evolution* 32:128–144.

Jobling, M.A.; Hurles, M.; and Tyler-Smith, C. 2004. *Human Evolutionary Genetics: Origins, Peoples and Disease*. New York: Garland Publishing.

Johns, G.C. and Avise, J.C. 1998. A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome b gene. *Molecular Biology and Evolution* 15(11):1481–1490.

Kang, S.; Kassam, N.; Gauthier, M.L.; O'Day, D.H. 2003. Post-mortem changes in calmodulin binding proteins in muscle and lung. *Forensic Science International* 131(2-3):140–147.

Kobilinsky, L. 1992. Recovery and stability of DNA in samples of forensic science significance. *Forensic Science Review* 4(1): 67–87.

Kocher, T.D.; Thomas, W.K.; Meyer, A.; Edwards, S.V.; Paabo, S.; Villablanca, F.X.; and Wilson, A.C. 1989. Dynamicas of mitochondrial DNA evolution in animals: amplification and sequencing with conserved priemrs. *Proceedings of the National Acadamey of Sciences USA* 86: 6196–6200.

Kumar, S. and Gadagkar, S. 2000. Efficiency of the neighbor-joining method in reconstructing deep and shallow evolutionary relationships in large phylogenies. *Journal of Molecular Evolution* 51:544–553.

Lehtonen, M. from Mitochondrial DNA sequence variation in patients with sensorineural hearing impairment and in the Finnish population. http://herkules.oulu.fi/isbn9514268490/html/index.html.

Lewin, B. 1997. *Genes VI.* New York: Oxford University Press and Cell Press.

Maddison, D.R. and Maddison, W.P. 2000. *MacClade 4: Analysis of Phylogeny and Character Evolution.* Sunderland, Mass., Sinauer Associates.

Maddison, W.P. 2000. Testing character correlation using pairwise comparisons on a phylogeny. *Journal of Theoretical Biology* 202:195–204.

Meyer, R.; Hofelein, C.; Luthy, J.; Candrian, U. 1995. Polymerase chain reaction-restriction fragment length polymorphism analysis: a simple method for species identification in food. *Journal of AOAC International* 78(6): 1542–1551.

Meyer, S.; Weiss, G.; and von Haeseler, A. 1999. Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics* 152:1103–1110.

Moraes, C.T.; Kenyon, L.; Hao, Huiling. 1999. Mechanisms of human mitochondrial DNA maintenance: the determining role of primary sequence and length over function. *Molecular Biology of the Cell* 10:3345–3356.

Naito, E.; Dewa, K.; Ymanouchi, H.; Kominami, R. 1992. Ribosomal ribonucleic acid (rRNA) gene typing for species identification. *Journal of Forensic Sciences* 37(2):396–403.

Nobrega, F.G. and Tzagoloff, A. 1980. Assembly of the mitochondrial membrane system: DNA sequence and organization of the cytochrome b gene in *Saccharomyces cervisiae*. *Journal of Biological Chemistry* 255(20):9828–9837.

Ono, T.; Miyaishi, S.; Yamamoto, Y.; Yoshitome, K.; Ishikawa, T.; Ishizu, H. 2001. Human identification from forensic materials by amplification of a human-specific sequence in the myoglobin gene. *Acta Medica Okayama* 55(3):175–184.

Parson, W.; Pegoraro, K.; Niederstatter, H.; Foger, M.; Steinlechner, M. 2000. Species identification by means of the cytochrome b gene. *International Journal of Legal Medicine* 114:23–28.

Parsons, T.J.; Muniec, D.S.; Sullivan, K.; Woodyatt, N.; Alliston-Greiner, R.; Wilson, M.R.; Berry, D.L.; Holland, K.A.; Weedn, V.W.; Gill, P.; Holland, M.M. 1997. A high observed substitution rate in the human mitochondrial DNA control region. *Nature Genetics* 15:363–368.

Rajapaksha, W.R.A.K.J.S.; Thilakaratne, I.D.S.I.P.; Chandrasiri, A.D.N.; Niroshan, T.D. 2002. Development of PCR assay for differentiation of some important wild animal meat of Sri Lanka. *Journal of Veterinary Medicine B* 49:322–324.

Roe, B.A.; Ma, D.P.; Wilson, R.K.; Wong, J.F. 1985. The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. *Journal of Biological Chemistry* 260(17):9759–9774.

Sarkioja, T.; Yla-Herttuala,S.; solakivi, T.; Nikkari, T.; Hirvonen, J. 1988. Stability of plasma total cholesterol, triglycerides, and apolipoproteins B and A-I during the early postmortem period. *Journal of Forensic Sciences* 33(6):1432–1438.

Scheffler, I.E. 1999. *Mitochondria*. New York, Wiley-Liss.

Shadel, G.S. and Clayton, D.A. 1997. Mitochondrial DNA maintenance in vertebrates. *Annual Review of Biochemistry* 66:409–435.

Stoneking, M.; Hedgecock, D.; Higuchi, R.G.; Vigilant, l.; Erlich, H.A. 1991. Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *American Journal of Human Genetics* 48:370–382.

Swofford, D.L. 1998. *PAUP* 4.0 – Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sunderland, Mass., Sinauer Associates.

Takahashi, K. and Nei, M. 2000. Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. *Molecular Biology and Evolution* 17(8):1251–1258.

Veltri, K.L.; Espiritu, M.; Singh, G. 1990. distinct genomic copy number in mitochondria of different mammalian organs. *Journal of Cellular Physiology* 143(1):160–164.

Wan, Q. and Fang, S. 2003. Application of species-specific polymerase chain reaction in the forensic identification of tiger species. *Forensic Science International.* 131:75–78.

Wetton, J.H.; Tsang, C.S.F.; Roney, C.A.; Spriggs, A.C. 2002. An extremely sensitive species-specific ARMS PCR test for the presence of tiger bone DNA. *Forensic Science International* 126:137–144.

Wilson, M.R.; Stoneking, M.; Holland, M.M.; DiZinno, J.A.; Budowle, B. 1993. guidelines for the use of mitochondrial DNA seqeuencing in forensic science. *Crime Laboratory Digest* 20(4):68–77.

Wilson, M.R.; DiZinno, J.A.; Polanskey, D.; Replogle, J.; Budowle, B. 1995. Validation of mitochondrial DNA sequencing for forensic casework analysis. *International Journal of Legal Medicine* 108(2):68–74.