J Forensic Sci, Jan. 2003, Vol. 48, No. 1
Paper ID JFS2002128_481
Available online at:www.astm.org

TECHNICAL NOTE

Wojciech Branicki,[1] *Ph.D.; Tomasz Kupiec,*[1] *M.S.; and Ryszard Pawlowski,*[1,2] *Ph.D.*

Validation of *Cytochrome b* Sequence Analysis as a Method of Species Identification*

ABSTRACT: One of the stages of dealing with biological material submitted to forensic laboratories is species identification. The aim of the present work was to validate and assess the possibility of applying sequence analysis of the region coding *cytochrome b* as a method of species identification in the field of forensic science. DNA originating from individuals from major phyla of vertebrates was isolated by the organic method from various specimens. Extracted DNA was subjected to PCR and direct cycle sequencing using a universal pair of primers. The validation process, performed according to TWGDAM recommendations, revealed that the technique is a very sensitive and reliable method of species identification allowing analysis of tiny amounts of material and also degraded material, and can be useful in the field of forensic genetics. The case example presented here, concerning the determination of species origin of biological evidence collected from fatal road accident, confirms that analysis can be carried out even when there is no reference sample, and the sequences obtained can be assessed through analysis of their similarity to sequences for *cytochrome b* present in DNA databases.

KEYWORDS: forensic science, *cytochrome b*, species identification, TWGDAM validation, GenBank, BLAST

Establishing species of origin is one of the basic aims of analyses applied in the process of identifying biological material in forensic laboratories. In court cases where the only material evidence is a trace of animal or plant origin, defining its species assumes a primary significance. Defining the species of origin is also becoming more and more significant in other fields such as the meat industry, fish processing and environmental protection (1). Because biological material that is sent to forensic genetics laboratories is frequently of low quality (trace degradation), the analytical method chosen must be highly sensitive and reliable. It seems that it would be more useful to study sequences of mitochondrial DNA (mtDNA) than nuclear DNA markers, because the large numbers of copies of mtDNA present in each cell significantly raise the sensitivity of the analysis. It is well known that in very old and highly degraded specimens, and also when investigating samples containing very small amounts of DNA such as hair shafts, mtDNA may be the only source for analysis (2,3). One of the regions of mtDNA used when establishing phylogenetic links between various species, and in species identification, is a fragment of the gene coding *cytochrome b* (*cytb*). It has been shown that this region can be amplified in various species of animals, using a single pair of universal primers in a PCR reaction under standard conditions (4,5). Recently Parson et al. have proposed application of the abundant DNA sequence data included in DNA databases for species identification of biological samples of unknown origin (6). This work presents results of the validation of the method of species identification by sequence analysis of the region coding *cytb*, applied according to TWGDAM recommendations (7). The method has been applied successfully to forensic casework and an example of its usefulness is also described.

Materials and Methods

Samples

DNA was extracted from various biological specimens—blood, saliva, soft tissues, animal hairs and bristles, birds' feathers, dried shed skin, old bones and additionally heated and processed meat. Specimens were subjected to overnight digestion in 10 mM Tris-HCl, pH 8.0/100 mM NaCl/10 mM EDTA/2% SDS/39 mM DTT/0.5 mg/mL proteinase K, followed by double phenol: chloroform: isoamyl alcohol extraction. DNA was then concentrated and cleaned using Microcon 100 (Millipore) concentrators. Negative controls were included in every extraction.

DNA was quantified using a fluorimetric dye, PicoGreen (Molecular Probes, Netherlands), and Fluoroskan Ascent FL apparatus (Labsystems, Finland).

DNA Amplification

The PCR reaction mixture was made up of 1.25U Taq polymerase (Promega), 200 μM dNTP (Promega), 3.1 μM of each of the primers, 2.0 μL 10× concentrated PCR buffer (Promega), 2mM MgCl2 (Promega) and usually 1–10 ng of DNA for 20 μL reactions. The following primers were used for amplification of the

[1] Institute of Forensic Research, Westerplatte 9, 31-033 Krakow, Poland.
[2] Department of Legal Medicine, Medical University of Gdansk, Sklodowskiej-Curie 3a, Gdansk, Poland.

* The project was partially supported by State Committee for Scientific Research in Poland no. 6P04 04216.

Received 9 April 2002; and in revised form 2 July 2002; accepted 4 July 2002; published 13 Nov. 2002.

cytb fragment (5):

Forward primer:
5′—CCATCCAACATCTCCGCATGATGAAA—3′
Reverse primer:
5′— CCCCTCAGAATGATATTTGGCCTCA—3′

Amplifications were performed in a Perkin Elmer 9700 thermocycler. The reaction conditions were settled as follows: 2 min at 94°C—initial denaturation, 32 or 36 cycles of 20 s at 94°C, 30 s at 51°C, 40 s at 72°C and final elongation step for 10 min at 72°C.

Negative control of DNA extraction and negative PCR control were subjected to all amplifications. The PCR products were evaluated by standard agarose gel electrophoresis in TBE buffer.

DNA Sequencing

Amplification products were purified using a Qiagen PCR Purification Kit (Qiagen, Germany) according to the manufacturer's directions. Sequencing reactions were performed in a Perkin Elmer 9700 thermocycler using the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems, UK). 5 × Sequencing Buffer (Applied Biosystems) was used for dilution of the BigDye Terminator Ready Reaction Premix. The reaction mixture consisted of 4 μL Ready Reaction Premix, 2 μL of 5 × sequencing buffer, 3.2 pmol of PCR primer and 2 to 30 ng of PCR products. Sequencing reactions were performed at 96°C for 10 s, 50°C for 5 s and 60°C for 4 min × 25 cycles. The sequencing products were electrokinetically injected on a 47 cm long capillary at 2.0 kV for 20 s followed by a 45 min run at 15 kV on an ABI310 machine. The sequencing data were analyzed with ABI Prism DNA sequencing analysis software ver. 3.0.

Sequence Similarity Search Using the BLAST Program

After sequence analysis, using Sequence navigator computer program ver. 2.1, all sequences were submitted as a query to the BLAST sequence similarity search. The BLAST computer program is widely used for searching protein and DNA databases for sequence similarities. Sequences where a similarity to the query is found are displayed according to the degree of sequence match. The BLAST is supplied by the NCBI on web site: http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/.

Validation Study

Sensitivity—Sensitivity of the method was assessed by analysis of consecutive dilutions of total DNA extracted from blood collected from representatives of two species: cat and pig. The dilutions were prepared in a range of 5 ng to 5 pg of DNA. PCR was performed using standard conditions described above, and 32 cycles. Problematic specimens containing usually low amounts of DNA as single hairs, bristles, bones and bird's feathers were also included in the study and in these cases the number of cycles was increased to 36.

Mixture Analysis—Possibility of detection of mixtures of DNA originating from different species was investigated by analysis of the DNA mixtures human/pig and dog/pig prepared in the range of: 1:1, 1:10, 1:50, 1:100, 10:1, 50:1, 100:1. This range of dilutions was chosen to explicitly show the fact that even very slight contamination by a specimen undergoing efficient amplification could affect the results of analysis.

Tissue Specificity—DNA extracted from blood, soft tissue, liver, heart, kidney and brain collected from one individual (domestic pig) was analyzed following the standard protocol.

The Effect of Various Substrates—Aliquots of 5 μL of cat blood were deposited in duplicates on jeans, suede, wood and wallboard, and analyzed after 5 days of storage in ambient temperature.

Exposure of Samples to Different Factors

Chemical Reagents—Aliquots of 5 μL of cat blood were deposited in duplicates on cotton treated by equal aliquots of gasoline, soap, 0.1 N NaOH, 5% acetic acid, 1 M NaCl, 0.5 M tannic acid, and analyzed after 5 days of storage in ambient temperature.

Exposure to Heat—Aliquots of 5 μL of cat blood were deposited in duplicates on cotton and placed in dishes floating in a water-bath settled at 37°C for periods of 24 h, 1 week, 1, 2, 3 and 4 months. After this time period samples were kept frozen in–80°C until extraction.

UV Irradiation—Blood aliquots were prepared as above and subjected to UV irradiation (Fotodyne transiluminator, UV 302 nm), from a distance of 15 cm for 1 h, 8 h and 24 h. After this time period samples were kept frozen in −80°C until extraction.

Results and Discussion

Primer Versatility

The study encompassed representatives of major phyla of vertebrates—fish, amphibians, reptiles, birds and mammals listed in Table 1; a total of 34 species were positively analyzed. The study confirmed the high degree of versatility of the applied universal primers (5). However, we noted evident differences in PCR efficiency depending on the species studied, which we assume could be influenced by point mutations occurring in the conservative region for which the primers were designed. A larger number of mismatches can lower the affinity of the primer for the DNA, which can be followed by lower efficiency of amplification. For example, the efficiency was significantly lower for the cow than for the pig. The typical PCR efficiency for humans is very high, which is beneficial, assuming that detection of human DNA is of primary importance in a forensic genetics investigation.

The research revealed that it was not possible to differentiate between some closely related species. The analyzed DNA fragment is not informative enough to enable unequivocal distinction between wild boar and domestic pig, mouflon sheep and domestic sheep, wolf and dog, or wild and domestic cat. These data hint that species belonging to the same genus could possess identical haplotypes, thus evading unambiguous species identification. In these cases, a more variable marker, such as hyper-variable parts of the control region of mtDNA could allow the differentiation of problematic samples. In some cases, analyses performed on a few individuals representing one species revealed unsubstantial variation, which cannot influence appropriate determination of species when using this methodology.

The significant advantage of this method is undoubtedly the possibility of identification of the species of origin of unknown samples, with no need for analysis of reference material. The fact that *cytb* region has been widely used for phylogenetic studies means that the GenBank contains sequences of this gene for a very broad range of animals. Use of the BLAST program allows one to define

TABLE 1—*Analyzed species.*

Species			
Latin Name	English Name		
	Mammals	Specimen	No. of Samples
Alces alces	Elk	Soft tissue	1
Bison bonasus	European bison	Blood	2
Bison bison	American bison	Bristles	1
Bos Taurus	Cow	Blood, heated meat	3
Canis familiaris	Dog	Blood, saliva, hairs	1
Canis lupus	Wolf	Soft tissue	3
Capreolus capreolus	Roe deer	Soft tissue, hairs	5
Equus caballus	Horse	Blood	1
Felis domesticus	Cat	Blood, saliva, hairs	4
Felis silvestris	Wild cat	Soft tissue	1
Felis lynx	Lynx	Soft tissue	2
Homo sapiens	Human	Blood, saliva, bones	10
Lepus europeus	Hare	Blood	2
Cercocebus aterrimus	Black mangabey	Hairs	1
Oryctolagus cuniculus	Rabbit	Blood, soft tissue	2
Ovis aries	Sheep	Blood, soft tissue	2
Ovis musimon	Mouflon sheep	Soft tissue	1
Pan troglodytes	Chimpanzee	Hairs	1
Rattus norvegicus	Rat	Blood, soft tissue	1
Sus scrofa	Wild boar	Soft tissue	2
Sus scrofa domesticus	Pig	Blood, soft tissue, heated meat	6
Ursus arctos	Brown bear	Soft tissue	1
	Birds		
Bubo bubo	Eagle owl	Soft tissue	1
Cairina moschata	Duck	Soft tissue	2
Ciconia ciconia	White stork	Feathers	1
Gallus domesticus	Chicken	Feathers	2
Meleagris gallopawo	Turkey	Soft tissue	1
Phasianus colchicus	Pheasant	Soft tissue	1
	Reptiles		
Anguis fragilis	Slow-worm	Soft tissue	1
Elaphe taeniura	Beauty snake	Dry shed skin	1
	Amphibians		
Bombina bombina	European fire-bellied toad	Soft tissue	1
Rana lessonae	Frog	Soft tissue	1
	Fishes		
Salmo salar	Salmon	Soft tissue, heated meat	1
Rutilus rutilus	Roach	Soft tissue	1

the degree of similarity of an evidence sequence with respect to all sequences accessible in the database (see case example). This approach was detailed by Parson, et al. (6). During the study we noted a lack of *cytb* DNA sequences in the GenBank for merely three species out of the 34 vertebrates investigated (*Cercocebus aterrimus*—Black mangabey, *Anguis fragilis*—Slow-worm, *Bombina bombina*—European fire-bellied toad).

Sensitivity and Mixture

The above-described fact of variations in PCR efficiency typical for distinct species influences the sensitivity of the method, which can thus vary slightly among different species. The analysis of consecutive dilutions of total DNA extracted from the blood of two species, cat and pig, revealed that the sensitivity is approximately 5 times higher for pig than for cat. Appropriate sequences were obtained for samples with 10 pg of pig template DNA input and 50 pg of cat template DNA input in the PCR reaction. In both cases, the sensitivity is comparable to that which is typical for commercial kits used for human identification purposes. Moreover, it is worth noting that the sensitivity was attained applying 32 PCR cycles, instead of the 36 usually applied to analysis of specimens containing tiny amounts of genetic material. One has to be aware, however, that the possibility of contamination rises when the number of PCR cycles is increased. The experiment performed in order to investigate the possibility of mixture detection confirmed that results strongly depend on the species comprising the analyzed mixture of biological material. In the case of human/pig DNA mixtures, non-interpretable sequences (nucleotides characteristic for both species) were detected for dilutions 1:1, 1:10, 1:50. The 1:100 sample gave a pure signal characteristic for pig, while the other samples expressed a signal typical for human DNA. In the case of the dog/pig DNA mixture; however, a pure signal issuing from dog DNA was not noted. The sequence characteristic for pig was ascertained in all samples, with two exceptions for dilutions 50:1 and 100:1, where the mixture was detected. We conclude that special care must be taken when surveying material containing very small amounts of DNA like hairs or bristles. This is because even minor contamination of a hair shaft by DNA originating from species characterized by very efficient amplification in comparison to the source of the evidence sample could lead to erroneous conclusions. We have had success concerning this issue by adhering to a decontamination protocol that allows for the removal of potential contaminant DNA from the surface of the hair shafts, described previously by Jehaes et al., which is based on prior treatment of the hair sample within an extraction buffer containing proteinase *K* but no DTT before the real extraction (8).

Tissue Specificity

All samples of DNA extracted from blood, soft tissue, liver, heart, kidney and brain of one representative of domestic pig were typeable and no trace of heteroplasmy was detected.

The Effect of Various Substrates

Aliquots of 5 μL of cat blood deposited on some substrates like denim, suede, wood, and wallboard were subjected to standard analysis. The obtained results are consistent with the abundance of available data in scientific papers on this matter. We noted negative results of analysis for blood deposited on wood, while Wilson et al. obtained 50% positive results from analysis for mitochondrial markers (2). The lack of PCR products observed in the case of suede was not unexpected, as the negative influence of leather on amplification has previously been described (9,2). The negative effect of inhibitors contained in denim on PCR reactions is also very well documented (10,11). Wilson et al. present data showing only 33% positive sequencing reactions from blood samples deposited on denim (2).

Exposure of Samples to Various Factors

Analysis of cat blood deposited on cotton and treated by various chemicals such as gasoline, soap, 0.1 N NaOH, 5% acetic acid, 1 *M*

NaCl, and 0.5 *M* tannic acid revealed a strong negative effect of tannic acid on PCR reactions. This influence can be explained by a previously described tendency of tannic acid to generate reactive oxygen species that cause DNA degradation, and the postulated direct DNA degradation caused by this reagent (12,13). The strong effects of tannic acid as well as other validation results of the *cytb* analysis were consistent with results of analogical validation performed on human HV1 and HV2 mitochondrial segments in our laboratory (unpublished data). Prepared validation bloodstains incubated at 37°C were fully typeable after 4 months of storage in dishes floating in a water-bath. Assuming damp conditions influencing the validation blood, this result confirms the strength of the method. The results of analysis of blood samples subjected to UV irradiation indicate a minor influence of this factor on reliable sequencing. Although UV light is a known factor affecting DNA, and is frequently used for decontamination (14), it is known that substances absorbing UV light adjacent to DNA can decrease the UV effect. Direct exposure to UV is also an important aspect of its effectiveness (15). The high reproducibility of the technique was additionally proven by positive results of analysis obtained for samples of heated and processed meat included in the study. We analyzed roasted and smoked meat, as well as processed sausages, and conclusively established the species of origin of all analyzed samples (Table 2). These results are fully consistent with conclusions of similar studies performed on heated specimens (16). Material typical for crime scenes and problematic specimens were also included in the study. Several bloodstains that were untypeable in nuclear STR markers gave positive results in *cytb* sequencing, and the species of origin was positively established. Good results were also obtained for analyzed bones, hairs, bristles, bird's feathers and shed snakeskin.

The fact that the validated method, based on the amplification of markers localized in mitochondrial DNA offers many advantages in analysis is due to high copy number of mtDNA per cell (3). The present study confirmed substantial stability of the marker against considerable influence of various chemical and physical factors. The obtained results are consistent with data presented in similar studies and authenticate the high degree of usefulness of this methodology for species determination performed in forensic genetics.

TABLE 2—*Results of validation study.*

Validation Samples	No. of Samples Analyzed	No. of Samples Successfully Typed
Mixtures*–blood/blood	14	14
Heated meat	4	4
Processed meat	2	2
Wood/blood	4	0
Jeans/blood	4	2
Suede/blood	4	0
Wallboard/blood	4	2
Gasoline/blood	2	2
Soap/blood	2	2
0.1 N NaOH/blood	2	2
5% Acetic acid/blood	2	2
1 M NaCl/blood	2	2
0.5 M Tannic acid/blood	2	0
Blood/UV 1h	2	2
Blood/UV 8h	2	2
Blood/UV 24h	2	1
Blood/37°C 24 h	2	2
Blood/37°C 1 week	2	2
Blood/37°C 1 month	2	2
Blood/37°C 2 months	2	2
Blood/37°C 3 months	2	2
Blood/37°C 4 months	2	2
Total	66	51

* - human/pig, dog/pig mixtures ranged from 1:1 to 1:100.

Case Report

The body of a woman bearing fatal injuries was found on a road. As a result of a police search, a suspected truck was identified. Detailed inspection of the truck showed tiny fragments of tissue on a mudguard. The truck driver was apprehended and interrogated as the main suspect of the deadly accident. The conclusion of a first study performed on the collected tissues in another laboratory stated that the evidence samples were too degraded to serve as a source of nuclear genetic data allowing determination of a potential match between the evidence and reference material. Our laboratory was asked to continue the identification process of the biological traces using more sensitive mitochondrial markers. Because STR analysis performed in the first laboratory was negative, it was even impossible to say whether the analyzed tissue sample was of human origin.

In the presented case, the evidence samples were analyzed as described above. The obtained nucleotide sequence was subjected to the BLAST search for identification purposes, based on the numerous animal *cytb* sequences present in the GenBank databases. Use of the BLAST program allowed us to define the degree of similarity of the sequence obtained with respect to all sequences accessible in databases. As a result of this search, we obtained a list of DNA sequences of animal origin that were significantly similar to the submitted sequence. In first place on the list was the *cytochrome b* sequence of the domestic cat (database accession number X82296.1), which was 99% identical to the sequence submitted as a query.

To prove our search results, reference samples taken from four different cats were analyzed and two haplotypes, varying in two positions (transitions A↔G), were identified. One of the ascertained haplotypes was identical with sequence X82296.1 and the second was identical with the haplotype identified in the evidence tissue.

Species of origin identification using sequence analysis of the *cytb* gene proved that the analyzed tissue sample was of nonhuman origin, and without any assumptions the species origin was assigned. This result unambiguously ruled out the suspicion based on the collected evidence (tissue) that the driver had been involved in the fatal road accident.

Conclusions

The detailed validation process of the *cytb* method performed in our laboratory according to TWGDAM guideline proved the high sensitivity, reliability and usefulness of the technique for forensic species identification. Using one pair of universal primers, a broad range of vertebrates can be investigated. However, unequal PCR efficiency typical for different species ought to be taken into account, especially when analyzing specimens containing a low amount of DNA. The variation typical for this region of mtDNA can be insufficient to enable unequivocal differentiation of two species belonging to the same genus. The low variability in *cytb* within one species does not interfere with positive species determination by sequence analysis. It was shown that positive results of amplification and sequencing can be obtained in samples containing heavily degraded or very low concentrations of DNA, such as animal hairs, birds'

feathers, dried skin, old bone samples or blood stains treated by various physical and chemical factors. Though we are aware that the validation sample sizes are low, our results are consistent with abundant literature data, and the strong effect of tannic acid was also observed during the validation process of human mitochondrial markers performed in our laboratory. The possibility of searching databases containing DNA sequences for *cytb* constitutes a significant advantage of this technique of species identification. It becomes possible to analyze biological material of unknown origin, even in a situation where a reference sample is not available.

Acknowledgments

Authors would like to kindly thank all those who provided biological samples used in the study. Special thanks in this matter are directed to the Executive of the Local Park and Zoological Garden in Krakow and personally to Bogumila and Wlodzimierz Jedrzejewski from the Institute of Mammals of the Polish Academy of Sciences in Bialowieza.

References

1. Raymond A, Forrest R, Carnegie PR. Identification of gourmet meat using FINS (Forensically Informative Nucleotide Sequencing). Biotechniques 1994;17:24–5.
2. Wilson MR, DiZinno JA, Polanskey D, Replogle J, Budowle B. Validation of mitochondrial DNA sequencing for forensic casework analysis. Int J Leg Med 1995;108:68–74.
3. Holland MM, Parsons TJ. Mitochondrial DNA sequence analysis—validation and use for forensic casework. Forensic Sci Rev 1999;11:21–48.
4. Kocher TD, Thomas WK, Meyer A, Edwards SV, Paabo S, Villablanca FX, et al. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc Natl Acad Sci 1989;86:6196–200.
5. Bartlett SE, Davidson WS. FINS (Forensically Informative Nucleotide Sequencing): a procedure for identifying the animal origin of biological specimens. Biotechniques 1992;12:408–11.
6. Parson W, Pegoraro K, Niederstaetter H, Foeger M, Steinlechner M. Vertebrate species identification by means of the *cytochrome b* gene. Int J Leg Med 2000;114:23–8.
7. Technical Working Group on DNA Analysis Methods: guidelines for the use of mitochondrial DNA sequencing in forensic science. Crime Lab Digest 1993;20:67–77.
8. Jehaes E, Gilissen A, Cassiman JJ, Decorte R. Evaluation of a decontamination protocol for hair shafts before mtDNA sequencing. Forensic Sci Int 1998;94:65–71.
9. Budowle B, Lindsey JA, DeCou JA, Koons BW, Giusti AM, Comey CT. Validation and population studies of the loci LDLR, GYPA, HBGG, D7S8, and Gc (PM loci) and HLA-DQα using a multiplex amplification and typing procedure. J Forensic Sci 1995;40:45–54.
10. Del Rio SA, Marino MA, Belgrader P. PCR-based human leukocyte antigen (HLA) DQ alpha typing of blood stained light and dark blue denim fabric. J Forensic Sci 1996;41:490–2.
11. Larkin A, Harbison S. An improved method for STR analysis of blood-stained denim. Int J Leg Med 1999;112:388–90.
12. Khan NS, Hadi SM. Structural features of tanic acid important for DNA degradation in the presence of Cu(II). Mutagenesis 1998;13:271–4.
13. Khan NS, Ahmad A, Hadi SM. Anti-oxidant, pro-oxidant properties of tanic acid and its binding to DNA. Chem Biol Interactions 2000;15: 177–89.
14. Prince AM, Andrus L. PCR: how to kill unwanted DNA. Biotechniques 1992;12:358.
15. Dieffenbach CW, Dveksler GS. PCR primer. A laboratory manual. Cold Spring Harbor Laboratory Press, 1995.
16. Meyer R, Hoefelein C, Luthy J, Candrian U. Polymerase chain reaction—restriction fragment length polymorphism analysis: a simple method for species identification in food. J AOAC Int 1995;78:1542–51.

Additional information and reprint requests:
Wojciech Branicki, Ph.D.
Institute of Forensic Research, Westerplatte 9
31-033 Krakow
Poland
E-mail: wbranic@ies.rakow.pl