Forensic Wildlife Parts and their Product Identification and Individualization Using DNA Barcoding

Raju Panday*, Dinesh K. Jha, Nirajan Thapa, Basant R. Pokharel and Nanda K. Aryal

National Forensic Science Laboratory (NFSL), Khumaltar, Lalitpur, Nepal

Abstract: Despite very strict wildlife protection acts in several countries and trading laws such as CITES, illegal trading and poaching of wildlife is still active mostly because of the use of their parts in Chinese medicinal products and ornaments. For successfully identifying wildlife and their parts, mitochondrial DNA sequencing is being used now-a-days. DNA bar coding using fragments of cytochrome b and cytochrome oxidase I genes are frequently utilized for mammalian species identification. This mini review describes some of the mtDNA sequences and STR markers used for mammalian species identification in the field of forensic wildlife along with the problems associated and their possible solutions.

Keywords: Cytochrome b, cytochrome oxidase, DNA bar coding, wildlife forensics.

INTRODUCTION

Earth is losing wildlife at an alarming rate and only few new species are being discovered each year. Illegal trade of wildlife parts and their products is worth more than 20 billion per year in black market [1]. In Nepal, every year millions worth of animal parts such as ivory, tiger skin and bone, bear bile, rhino horn and other wildlife materials are illegally transported across the borders mostly via Tibet. This kind of illegal trade creates a serious threat to the global as well as to the local diversity either directly or indirectly. Consequently the role of international organizations such as WWF (World Wildlife Fund), IUCN (International Union for Conservation of Nature) and CITIES (Convention on International Trade in Endangered Species of Wild Flora and Fauna) becomes more valuable than ever in protecting wildlife and their habitat. Habitat destruction causes increase of wildlife poaching due to their ever increasing vulnerability of being caught by humans, so that wildlife trade should also be cautiously examined and monitored for any critical illegal transactions [2].

WILDLIFE CONSERVATION

CITES regulates the global trade of wildlife within its 180 signatory countries [3]. It operates by listing wildlife (approximately 5,600 animal and 30,000 plant species are currently listed by CITES) in one of its three appendices (Appendix I, II and III). Since it is a trading treaty, it allows free and controlled trade of wildlife parts and their products from non-endangered wildlife (flora and fauna both), whose survival is not yet threatened but the species registered in Appendix I (630 fauna and 301 fauna) [3] are banned from such trade, except for scientific research and other specific purposes. Although CITES is legally binding to the signatory countries, it does not replace national laws. Problems arise when species of CITIES appendix I cannot be verified by the laboratory when confiscated and sent for identification by the forest officers. Most of the time, domestic animal parts are confused with parts from protected and endangered animal species, as it is difficult to identify them through morphological and anatomical studies only. The difficulties that arise after the capture of wildlife parts or their products are: (i) determining the species that was killed with the help of a bloodstain only, (ii) when the sample is in altered or powdered form, (iii) identification up to family of animal but not to the specific species, and (iv) sometimes not even a clue, regarding the wildlife from which the sample originated, can be deciphered from basic morphological or anatomical test procedure [4].

Species identification becomes more problematic because most of the times it’s not the whole animal part but rather only a small piece or some modified part such as musk cone, rhinoceors horn, elephant ivory, elephant tail hair, tiger skin and bone that is traded [1]. This leads to complexity in precisely identifying the concerned protected wildlife species. The suspected animal is then either excluded or not excluded or a family of the animal is discerned for the species but not the species itself. This limitation can be remedied by using molecular tools and techniques for wildlife parts and product identification. Newer techniques such as DNA sequencing can help in identifying the species from which the product originated with great accuracy and reliability.

SPECIES IDENTIFICATION

For the purpose of species identification, although for most of the intact specimens the morphological observations may be enough, but in samples which contain only some part of an animal or their modified products (e.g. powder), DNA based methods are the methods of choice. In the text that follows, we first describe the limitations of conventional
methods of identification and then provide details of the DNA based methods.

Conventional Methods and Their Limitations

When gross morphological characteristics, such as whole skin or skeleton are present in a sample, then the general morphological, anatomical, microscopic and osteological analysis can identify the species with certain confidence [5]. When skin samples are present, hair analysis and comparison with reference sample can be fruitful in identifying the species of origin for that sample, but it requires a database of known/reference hairs and a high level of expertise in this field. Sometimes, antigen antibody reactions such as Ouchterlony test can also be used to identify animal products such as tissue samples. However, it has several disadvantages; firstly there is difficulty in getting the specific antibody as not all animal species specific antibodies have been produced [6, 7]. Furthermore the test can sometimes produce cross-reactions in non-target species [8]. Therefore, it has been suggested that the test should only be considered as a presumptive test in the forensic environment [9].

More reliable results can be obtained by DNA-based tests such as Random Amplified Polymorphic DNA (RAPD) [10, 11], Amplified Fragment Length Polymorphism (AFLP) [12, 13] and Random Fragment Length Polymorphism (RFLP) [14-16]. These allow a specific DNA profile to be produced for an individual animal. However, they suffer from the fact that they are not always reproducible and in the long run the preparation of a reference database is also not possible [17, 18]. These tests also cannot reliably separate mixture samples and in addition, within the same species some variations exist between samples which cannot be separated. In these tests, the band patterns obtained are influenced by many factors such as the quality of template, buffer, dNTPs, and primer concentration [19].

Random amplified polymorphic DNA (RAPD) markers can be amplified relatively easily using standard protocols. There is no requirement for any prior genetic information about the species in question and the method is relatively simple and inexpensive [20]. A more severe weakness of RAPD is that the low stringency PCR required in the procedure, results in a high genotyping error rate and in lower genotyping reproducibility compared to single locus markers [21]. In AFLP and RFLP, the tests can be used to generate species-specific band patterns even if no prior gene sequence knowledge is known [22, 23]. Several restriction enzymes are required for generation of specific band pattern for a species [24]. These methods cannot separate mixtures having the same restriction patterns and sometimes cannot separate closely related species [25, 26].

DNA Barcoding and Mitochondrial DNA Sequencing

The above mentioned limitations in traditional DNA-based methods for identification, led to the development of a more robust concept of DNA bar coding around the world in recent years. DNA barcode, is a short segment of a gene, usually taken from mitochondrial DNA region, which is used to identify wildlife species and their parts.

DNA barcoding uses the regions of DNA that show considerable interspecies variation and no or little intraspecies variation. Table 1 briefly summarizes the currently followed DNA based tests that are used for wildlife species identification and individualization. Among them the most popular DNA based method used in wildlife identification has been the mitochondrial DNA sequencing [1, 27, 28]. Human mtDNA is 16,569 bp in length which contains 13 genes, 2 rRNAs and 22 tRNAs (Fig. 1). The mtDNA is usually larger in plants (180 to 700 kbp) than in animals (14 to 20 kbp) due to the presence of a number of introns and pseudo genes [29]. In mammals, the two mtDNA strands are designated as H and L strands. H strand is heavier than L strand since it is rich in GC content. It separates in CsCl (Cesium chloride) gradient separation [30]. The reference sequence used is the 5'-3' L strand [30].

Mitochondrial DNA has some features, listed below, that make it a useful tool for species identification:

- Each cell has 100s of copies of mtDNA [31, 32].
- Mitochondrial DNA is better protected from degradation due to its own rigid membrane which is high in protein content. It can even be found in highly degraded samples that do not contain much nuclear DNA (e.g. hair, bone) [33].
- There is no proofreading activity during mtDNA replication so there is greater chance of mutation or change in DNA sequence than in the nuclear DNA [34, 35].

Table 1. Summary of different DNA based genetic tests currently available.

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Genetic Test</th>
<th>Species Identification</th>
<th>Individual Identification</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>mtDNA sequencing</td>
<td>Yes</td>
<td>No</td>
<td>Suitable for degraded DNA samples</td>
<td>Heteroplasmy</td>
</tr>
<tr>
<td>2.</td>
<td>STR (Short Tandem Repeat) typing</td>
<td>No</td>
<td>Yes</td>
<td>Highly informative markers for many applications</td>
<td>Allele dropout in degraded DNA samples</td>
</tr>
<tr>
<td>3.</td>
<td>Species specific primers</td>
<td>Yes</td>
<td>No</td>
<td>Capable of rapid screening</td>
<td>Knowledge of species boundaries required</td>
</tr>
<tr>
<td>4.</td>
<td>Pyro-sequencing</td>
<td>Yes</td>
<td>Yes</td>
<td>Rapid high throughput genotyping of STRs and SNPs</td>
<td>Only short (10 to 500 bp) DNA fragments can be sequenced</td>
</tr>
<tr>
<td>5.</td>
<td>SNPs</td>
<td>Yes</td>
<td>Yes</td>
<td>Highly reproducible</td>
<td>About five times more loci required than STRs</td>
</tr>
</tbody>
</table>
Mitochondrial DNA is maternally inherited so all the maternal lineages will have the same mtDNA sequence [36, 37]. Species identification most commonly uses sequencing of the cytochrome b (Cytb) gene, 12S and 16S rRNA segment, control region (D loop) and cytochrome c oxidase subunit I (COI) gene in animals; rbcL and matK (plastid genes) in plants. In plants, mtDNA is not used for species identification because of the low rate of sequence mutation which cannot provide species level resolution. In fungi, ITS (internal transcribed spacer regions) of 12S rRNA is used for the purpose of species identification. Some of the DNA barcoding genes (regions) are described below.

**Cytochrome b (Cyt b) Gene**

It is an 1140 base pair long gene coding for a 380 amino acid protein which is involved in electron transport and that makes complex III of mitochondrial oxidative phosphorylation system in humans [5, 38, 39]. Its length typically varies within species and is used for taxonomic purpose to establish phylogenetic relationship of wildlife species [29]. The mostly used region of Cyt b is a 385 bp long segment that has over 8,000 sequences available in Genbank for vertebrates (Table 2) [40]. Some experts [41] have used a 402 bp fragment to identify animal species, which showed less than 3% variation within any species. Some researchers [42] have even used 900 bp segment of Cytb for species identification (Table 2).

Both variable and conserved regions are present in cytochrome b. Since the variable regions are not subject to functional constraints, they seem to be positioned within the coding regions for transmembrane domains or for the amino and carboxy terminal ends [40]. In one study, identification was not possible with the COI gene since the reference sequence was not available in Genbank and BOLD for the investigated species [43]. In some species, therefore, Cyt b is the best represented gene in Genbank which has a higher ability to separate species when compared with COI [44].

**12S and 16S rDNA Sequence**

Mitochondrial 12S and 16S rDNA have been mostly applied in studies of higher categorical levels such as phyla and family respectively [21]. These sequences have been useful for inference of moderate to long divergence times [45]. The 16S rRNA gene encodes for the mitochondrial large ribosomal subunit (mt LSU) in animals and has been widely used in order to explore phylogenetic relationships in arthropods at most phylogenetic levels, familial level, genus level and below [46].

Molecular identification and phylogenetic analysis have been performed using 12S rDNA in Indian muntjac (Muntiacus muntjak) [47]. 16S rRNA partial sequence has been utilized for the identification of crocodile species in India [46]. In another study involving mitochondrial rRNA genes of Chinese Antelopes, average sequence divergence values for 12S rRNA and 16S rRNA genes were found to be 6.3% and 9.9% respectively [48]. The 16S rRNA...
sequence has also been shown to be superior to COI gene for amphibian species identification [49]. They have a greater sequence variation than Cyt b gene for a species and the process is robust and well validated [50].

**Cytochrome c Oxidase I (COI) Gene**

Barcode for life database (BOLD) by the Smithsonian Institute has been using COI as the standard loci for species identification [51]. COI gene segment from the sample is first sequenced, and is then compared with the known published sequence database, consortium for the Barcode of Life (CBOL) (Fig. 2) [52]. It is possible to identify unknown samples once the sequences from enough species are accumulated on the CBOL database (Table 2). A 648 segment of the COI gene (also called DNA barcode), which has the capability to distinguish 95% of the current species in the world, is sequenced [52]. It has advantages over Cyt b in that there is a broad range of primers applicable across a wide range of taxa and genetic changes occur more slowly than in Cyt b [53, 54]. The COI gene is the current focus of a worldwide effort to characterize all species through a single genetic entity. This will cause reliance on CBOL and the test will become limited to the identification of species whose sequences have been uploaded onto CBOL database. Currently, CBOL has in its database 2,891,971 sequences from 192,480 species of animals, plants and fungi [55].

It should be noted that universal primers used for COI gene do not amplify consistently in all species in a genus/family/order due to possible point mutations at the primer binding site that results in incomplete amplification [9]. These sites could be further investigated for use as a possible SNP based identification test. Another drawback of using COI gene is the lack of an extensive database across all the genus/family/order of animals. For example, only about 16% of the total mammalian species sequences have been added to the CBOL database which needs to be extensively expanded in order to accommodate remaining species.

**Control Region (CR) Sequence**

The control region (CR) or D loop in humans has two hyper-variable regions (HV1 and HV2) with a central region in between (Fig. 1). Proteins, responsible for replication and transcription, bind to the D loop [56-58]. In animals, however, CR is composed of three domains including ETAS (extended termination associated sequences) domain, central domain and CSB (conserved sequence block) domain [21]. The CR region of tiger has been divided into five parts: hypervariable region I (HVI), repeat sequence 2 (RS 2), conserved control region (CCR), repeat sequence 3 (RS 3), and hypervariable region II (HVII) [59].

The CR spans between the tRNA^Phe^ and the tRNA^Pro^ genes and includes the transcription initiation sites for both strands [60, 61]. The CR is usually considered to be the fastest-evolving region of the genome and for that reason it is broadly used to perform intra- and inter-specific phylogenetic studies [62].

One such study involved analysis of the 403 bp fragment of mitochondrial DNA CR for the genetic diversity and population structure of Chinese water deer, which revealed 18 different haplotypes in 40 samples [63].

It is important in closely related lineages, but not much suitable for species identification because of back and parallel mutations observed. It is used for species identification in animals and for individual identification in humans using the HV1 and HV2 regions of D loop [64].

**Individual Identification Using Nuclear DNA**

Mitochondrial DNA analysis only gives the identification of species but cannot discern the individuality among the same species being identified. Identification of an individual based on its unique genetic profile can be used to monitor the number of animals entering commercial markets, even if they are sold as meat or highly processed products. Individual identification, which is most commonly achieved using STR profiling of nuclear DNA instead of mtDNA, may be required in cases where a DNA match is being investigated between an illegal kill (e.g. remains) and evidence obtained from a poacher (e.g. skin or meat) [65]. However, not much work has been carried out in the field of non-human STR typing. Few papers such as that of Singh *et al.* [66] discovered a DNA marker called Ple46 in four species of big cats (*Panthera* species) in India where different animals had different lengths of the microsatellite marker (Table 2). The repeat sequence was CA bases; domestic cat had 10 (CA) repeats, lion had 22 repeats, leopard had 14-15 repeats and tiger had 7-8 repeats while the heterozygosity level for Ple46 marker was high (>75%) [66].

In another study, more than 29 STR markers were used in different combinations for individual identification of domestic dog [84]. While performing individual identification with STRs, generating a Random Match Probability (RMP) value is also necessary if there is a match between the evidence and reference sample. For this reason, STR DNA databases are required which, for wild endangered species, are very scarce. Some existing relevant databases are that of African Elephants [85], Mule deer in Alberta [79], Wild boars in Italy [81] and Tigers in India [86]. STR technology has been used for the prosecution of humans accused of animal cruelty [87] and theft of cattle [28], birds of prey [88] in forensic wildlife crime cases.

However, there is a problem with the current STR profiling being used for wildlife species individualization. Even when much sequence data is available, it does not have much value in forensic cases because most of the literature is flooded with dinucleotide STR markers which are not ideal in forensic science because of the generation of stutter products [66]. At least tetra nucleotide or larger STR markers need to be identified in order to be of any use in forensic identification. Therefore, much of the QC, QA and method validation must be conducted before using STR markers for a species identification in forensic caseworks.

**PROBLEMS ASSOCIATED WITH mtDNA MARKERS AND THEIR POSSIBLE REMEDIES**

Mixed samples are the most often experienced problems with the mtDNA markers. In case of a contaminated sample with DNA from another species, the results obtained will be
unreliable because of the use of universal primers that bind to DNA of both species [56] which is very difficult to mitigate. In an approach to identify species in mixed samples, analysis of CR inter-specific length polymorphism enabled a very simple and accurate mammalian species identification using a single pair of universal primers [89].

Another problem associated with the current generation sequencing of mtDNA genes is that if a hybrid species is protected by law and it is produced by using a non-protected maternal animal species, then the mtDNA analysis will result in the profile of the maternal non protected species [90]. This kind of problems can be very serious when dealing with the hybrid protected species identification process [91].

Heteroplasmym is yet another issue that can create problems when dealing with identification [92, 93]. It is the situation where an individual has different copies of mtDNA due to mutation [94]. This does not affect species identification but can cause difficulty when looking within the species [95].

Specimen is usually killed using alcohol or freezing.

Small tissue section is removed.

Chelax100 or silica membrane based methods are used for DNA extraction.

Real Time PCR is used for this purpose.

Primers for COI 648 fragment are used along with Trehalose as PCR enhancer and DNA polymerase such as Z Taq is used for PCR.

Pre-cast agarose gels are used.

BigDye cycle sequencing performed.

Sequencers such as ABI 3100/3100XL/3500 series are used.

Done using BLAST in Genbank, EMBL databank.

VALIDATION OF METHODS FOR FORENSIC SCIENCE

Validation is the process of demonstrating that a laboratory process is robust, reliable, and reproducible in the hands of the personnel who performs the test [96]. Commercially available technology and the kits that are used in DNA analysis in forensic cases are validated by the manufacturer before the product is released to the market. Therefore, the forensic laboratory has to carry out only some internal verification work so that it can meet the accepted validation requirements [96].

Scientific Working Group on DNA Analysis Methods (SWGDAM) and the ISO17025 or IEC17025 laboratory accreditation are the most common standards employed in the forensic field. Both standards were established in the late 1980s under FBI to help forensic scientists in forensic examination. They include testing of technology and verification for the reproducibility against standard samples, samples in more complex matrices, mixed samples and
samples exposed to a variety of environmental conditions. Rules are guided to ensure that reliable results are obtained with the actual forensic samples in the laboratory [97].

Forensic scientists need to be well aware of the legal system and the rules and expectations, as an expert witness, if they are to bring wildlife forensics into the court system. This will require the use of validated laboratory methods and accredited laboratory to ensure reproducibility and admissibility of the test upon which the court can rely.

CONCLUSION

With an ever increasing wildlife crime and no sign of decreasing, there is an urgent need to develop forensic processes for wildlife identification that meet international standards. Each and every wildlife species should be identifiable. There are many standard protocols for species identification and currently mitochondrial DNA based identifications are mostly being followed around the world. With BOLD making COI gene as a standard for species identification in wildlife cases, there exist validated techniques that can also be legally used. Supplementing DNA sequencing results in identification, with STR markers for individualization, can help build confidence in results by the courts. Thus the research concepts used in human identification can also be used in wildlife crime and its eradication.

Table 2. Summary of literature for species and individual identification in animals using different genes and their fragments.

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Sample and its Species</th>
<th>DNA Marker Used</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Skin; Crocodile</td>
<td>COI (645bp)</td>
<td>[67]</td>
</tr>
<tr>
<td>2</td>
<td>Meat, carcass; Reedbuck</td>
<td>COI (650bp)</td>
<td>[68]</td>
</tr>
<tr>
<td>3</td>
<td>Hair, leather; Asiatic black bear</td>
<td>COI (708bp)</td>
<td>[69]</td>
</tr>
<tr>
<td>4</td>
<td>Feathers; Scarlet macaw</td>
<td>COI (648bp)</td>
<td>[70]</td>
</tr>
<tr>
<td>5</td>
<td>Wool; Tibetan Antelope</td>
<td>Cyt b (311bp)</td>
<td>[71]</td>
</tr>
<tr>
<td>6</td>
<td>Tissue; Cobra snake</td>
<td>Cyt b (472bp)</td>
<td>[72]</td>
</tr>
<tr>
<td>7</td>
<td>Ivory; Elephant</td>
<td>Cyt b (357bp)</td>
<td>[73]</td>
</tr>
<tr>
<td>8</td>
<td>Caviar; Sturgeons and paddlefish</td>
<td>Cyt b (850bp)</td>
<td>[74]</td>
</tr>
<tr>
<td>9</td>
<td>Meat; Whale species</td>
<td>Cyt b (400bp)</td>
<td>[75]</td>
</tr>
<tr>
<td>10</td>
<td>Bile crystals; Asiatic black bear</td>
<td>Cyt b (175bp)</td>
<td>[76]</td>
</tr>
<tr>
<td>11</td>
<td>Meat, hair; Roe deer</td>
<td>Cyt b (900bp)</td>
<td>[77]</td>
</tr>
<tr>
<td>12</td>
<td>Meat; Indian peafowl</td>
<td>Cyt b (472bp)</td>
<td>[78]</td>
</tr>
<tr>
<td>13</td>
<td>Horn, Rhinoceros</td>
<td>Cyt b (402bp)</td>
<td>[79]</td>
</tr>
<tr>
<td>14</td>
<td>Blood; Tiger</td>
<td>STR markers</td>
<td>[80]</td>
</tr>
<tr>
<td>15</td>
<td>Blood; Eurasian badger</td>
<td>STR markers</td>
<td>[81]</td>
</tr>
<tr>
<td>16</td>
<td>Solid tissue; Mule deer</td>
<td>STR markers</td>
<td>[82]</td>
</tr>
<tr>
<td>17</td>
<td>Tail tissue; Sea horse</td>
<td>STR markers</td>
<td>[83]</td>
</tr>
<tr>
<td>18</td>
<td>Solid tissue; Wild boar</td>
<td>STR markers</td>
<td>[84]</td>
</tr>
<tr>
<td>19</td>
<td>Blood, horn; Rhinoceros</td>
<td>STR markers</td>
<td>[85]</td>
</tr>
<tr>
<td>20</td>
<td>Tusk; African Elephant</td>
<td>STR markers</td>
<td>[86]</td>
</tr>
</tbody>
</table>

CONFLICT OF INTEREST

The author confirms that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES


Calvo, J. H.; R. Osta; Zaragoza, P. Species-specific amplification for detection of bovine, ovine and caprine cheese. Milchwissenschaft, 2002, 57, 444-446.


