Best Practice Manual for the Application of Molecular Methods for the Forensic Examination of Non-Human Biological Traces

ENFSI-BPM-APS-01

Version 01 - November 2015

With the financial support of the Prevention of and Fight against Crime Programme
European Commission - Directorate -General Home Affairs

Background

This Best Practice Manual (BPM) belongs to a series of 10 BPMs issued by the European Network of Forensic Science Institutes (ENFSI) in November 2015. The series covers the following forensic disciplines:

1. Forensic Examination of Digital Technology
2. Forensic Examination of Handwriting
3. Chemographic Methods in Gunshot Residue Analysis
4. Road Accident Reconstruction
5. Microscopic Examination and Comparison of Human and Animal Hair
6. Fingerprint Examination
7. DNA Pattern Recognition and Comparison
8. Application of Molecular Methods for the Forensic Examination of Non-Human Biological Traces
9. Forensic Recovery, Identification and Analysis of Explosives Traces
10. Forensic Investigation of Fire Scenes which have resulted in Fatalities*
11. Forensic Investigation of Fire Scenes which involve the Clandestine Manufacture of Improvised or Homemade Explosive Devices*
12. Forensic Investigation of Fire Scenes which Involve the Clandestine Manufacture of Illicit Synthetic Drugs*

* *The three specific areas on Forensic Investigation of Fire Scenes (numbers 10 -12) were combined into one BPM 'Investigation of Fire Scenes'.*

In the years 2014 and 2015, so-called Activity Teams have - in parallel - developed the 10 BPMs. The activities were performed within the project 'Towards European Forensic Standardisation through Best Practice Manuals (TEFSBPM)' and co-ordinated by the ENFSI Quality and Competence Committee. The realisation of the BPMs was supported by the Prevention of and Fight against Crime Programme of the European Commission – Directorate General Home Affairs (code: PROJECT HOME/2012/ISEC/MO/4000004278). The core project concept was that the BPMs will enhance the quality of the forensic services available to law enforcement and justice across Europe and thereby encourage forensic standardisation and cross-border cooperation between countries.

ENFSI expects that the issuing of this series will stimulate the improvement of already existing BPMs as well as the creation of new BPMs on disciplines that are not covered yet.

Acknowledgements
Luc Bourguignon (National Institute of Criminalistics and Criminology - Belgium)
José Carlos Cordero Pérez (Criminalistic Service of the Civil Guard - Spain)
Stijn Desmyter (National Institute of Criminalistics and Criminology - Belgium)
Balazs Egyed (Synlab GenoID - Hungary)
Andreas Hellmann (Federal Criminal Police Office - Germany)
Irene Kuiper (Netherlands Forensic Institute - The Netherlands)
Zsolt Padar (Synlab GenoID - Hungary)
Enrique Javier Santillana Reyes (Criminalistic Service of the Civil Guard - Spain)
Christina Staginnus (State Criminal Police Office Rheinland-Pfalz - Germany)

Official language
The text may be translated into other languages as required. The English language version remains the definitive version.

Further information
For further information about this publication, contact the ENFSI Secretariat. Please check the website of ENFSI (www.enfsi.eu) for update information.

Best Practice Manual for the Application of Molecular Methods for the Forensic Examination of Non-Human Biological Traces

CONTENTS

1. AIMS

This Best Practice Manual (BPM) aims to provide a framework of procedures, quality principles, training processes and approaches to the forensic examination. This BPM can be used by Member laboratories of ENFSI and other forensic science laboratories to establish and maintain working practices in the field of forensic molecular genetic analysis of non-human biological traces (NHBT) that will deliver reliable results, maximize the quality of the information obtained and produce robust evidence. The use of consistent methodology and the production of more comparable results will facilitate interchange of data between laboratories.

The diversity in the type of traces within this field, the huge variability of population dynamics and the genetics of these traces as well as the development of new molecular tools and techniques (DNA/RNA), makes it undesirable to formulate best practices. The term BPM is used to reflect the scientifically accepted practices at the time of creating. Despite its implicit suggestion that alternative, equivalent Practice Manuals are excluded at beforehand, in this series of ENFSI Practice Manuals the term BPM has been maintained for reasons of continuity and recognition.

Some recommendations for the interpretation of the results of the analyses and the expression of the criminalistic value of these will be provided. These recommendations apply to NHBT and cover the advantages and disadvantages of their use in forensic casework.
It should be emphasized that this manual is made to encourage a systematic approach and high forensic quality without limiting the use of novel DNA/RNA tools or uncommon forensic traces. For that purpose the manual gives minimum requirements for effective validation and application rather than best practices. Recommendations are provided in order to align methodologies between forensic laboratories, allowing the exchange of knowledge, expertise and databases.

2. SCOPE

This BPM is aimed at experts in the field and assumes prior knowledge in the discipline. It is not a standard operating procedure and addresses the requirements of the judicial systems in general terms only.

It serves as a framework for the minimum requirements and recommendations for the molecular genetic analysis of NHBT in forensic casework. Forensic casework starts at the crime scene where the collection and securing of traces should be done in a methodical manner to ensure as many analytical methods as possible can be used. The next step in the process is examining and describing traces to formulate an analytical plan. The analytical plan is dependent on the forensic question, available technologies, preservation of trace evidence and the criminalistic value of the analytical results.
This framework includes the securing of NHBT at the crime scene and from pieces of evidence, their examination and subsequent analysis using molecular tools, the interpretation of data and formulation of a forensic report. Additional tools and methods such as immunology and isotope analysis are applicable but are not addressed in this manual.
In contrast to investigation of human DNA-traces, there are no restrictions concerning DNA loci and privacy of genetic information from animals, plants and microorganisms. A wide range of technologies can be applied due to the wide spectrum of genetic information (DNA, RNA) and organization (e.g. inheritance, reproduction strategies) which may be encountered.
When DNA- or RNA-markers are recommended, this is only done in order to facilitate the

exchange of data and the alignment of methods and databases, without giving more value to these markers or devaluing other markers that meet the same minimum requirements. Tools such as DNA/RNA-markers can be highly variable within different populations and new tools are developed on a regular basis.
For example, individual typing using nuclear DNA-markers like STRs in both plants and animals requires the same instrumentation, reagents and technical expertise, while the final results can have very different criminalistic value if profiles are generated from plants which can easily propagate clonally (like Cannabis) or when profiles are generated from traces of a dog.
For certain types of species and biological traces additional recommendations for the interpretation of results are provided to highlight the limitations and restrictions.

3. DEFINITIONS AND TERMS

For the purposes of this Best Practice Manual (BPM), the relevant terms and definitions given in ENFSI documents, the ILAC G19 “Modules in Forensic science Process”, as in standards like ISO 9000, ISO 17000 and 17020 apply.

16S/12S rRNA gene	genes located on the genome encoding 16S/12S rRNA
-A peaks (or double peak)	PCR artefact that is one nucleotide shorter than the amplified PCR product (allele)
allele	possible status of a gene in a certain locus in the genome
allelic drop out	preferential amplification of one out of two alleles due to mutations in the primer binding regions and/or low amount and degradation events of target DNA
amplicon	defined DNA stretch amplified by PCR
blind trial	internal quality measure using reference samples to ensure the correctness of the results
BOLD	Barcode of Life Data Systems; nucleotide sequence database devoted to taxonomic studies and online platform for DNA sequence analysis à http://www.boldsystems.org/
bp	base pair
carry-over	unintentional transfer of biological material (including DNA) between the analysed samples
CBOL	Consortium for the Barcode of Life
CE	capillary electrophoresis; analytical method allowing the separation of DNA molecules according to their molecular weight
CE	collaborative exercise
chemotyping	determination of a biochemical phenotype (chemotype) on the DNA/RNA level
COI	cytochrome oxidase I; gene located on the mtDNA
cpDNA	chloroplast DNA; portion of the genome located in the chloroplast
cytb	cytochrome b; gene located on the mtDNA
D-Loop	displacement loop; noncoding structure occurring in mitochondrial circular DNA molecules
drop in	additional allele in a genetic profile due to PCR artefact (e.g. stutter) or unintentional contamination

family	taxonomical classification, family comes between order and genus
GENBANK	open access nucleotide sequence database produced and maintained by the National Center for Biotechnology Information (NCBI) à http://www.ncbi.nlm.nih.gov/
genetic profile	combination of alleles of multiple genetic markers in an individual being
genus	taxonomical classification, genus comes between species and family
GroEL	gene found in a large number of bacteria required for the proper folding of many proteins
INDEL	insertion or deletion of bases (1 – 50 nucleotides)
ITS	internal transcribed spacer; sequences located on the ncDNA
Locus	certain position in the genome
LSU	Large subunit 25-28S ribosomal RNA
Marker	synonym to locus, carrying one or multiple alleles
match	identical alleles in all markers investigated, the DNA profile obtained from two samples are indistinguishable
matK	gene located on the cpDNA
mismatch	divergence in alleles of two genetic profiles compared
mtDNA	mitochondrial DNA; portion of the genome located in mitochondria
ncDNA	nuclear DNA; genome located in the nucleus
NGS	next generation sequencing; high-throughput sequencing technologies for DNA molecules
NHBT	non-human biological traces; biological evidence material of non-human origin
PCR	polymerase chain reaction; molecular technique allowing the amplification of a defined DNA molecule
Primer	oligonucleotide used as starter molecule in a PCR reaction
PT	Proficiency Test
pull-up peaks	false signal due to the detection system/electrophoresis
rbCL	gene located on the cpDNA
RISA	Ribosomal Intergenic Spacer Analysis
SNP	single nucleotide polymorphism; variation in a single nucleotide position of a DNA sequence
species	taxonomical classification of a group of organisms capable of cross-breeding (and for animals capable of producing fertile offspring)
spikes	signals not derived from DNA, they generally appear in all detection channels of the CE and are sharper than regular peaks
STR	short tandem repeat; class of short repetitive sequences revealing variable numbers of their basic repeat unit in a tandem array
stutter peaks	PCR artefact that is usually one repeat shorter (or longer) than the amplified DNA product / allele
taxonomic typing	determination of a taxon on the DNA/RNA level
tRFLP	Terminal Restriction Fragment Length Polymorphism
trnL/F/H	gene located on the cpDNA
psbA	gene located on the cpDNA

4. RESOURCES

4.1. Personnel

Personnel should be educated in accordance with national rules and regulations. They should be formally educated or trained in the application of current molecular methods. This includes also knowledge of what is commonly accepted in the field and awareness of emerging fields. Experts writing reports should be aware of differences of the area of expertise to the nearest related fields (e.g. like human DNA) due to ecological aspects and genetic background of the species investigated (see Chapter 6 and 12) and they should be aware of new developments in related fields.

4.2. Equipment

The equipment should fulfil the specific requirements of the field. In general, instruments for DNA extraction, amplification, separation and detection are required. The specific software should be chosen according to the applied method.

4.3. Reference materials

Commercial reference materials for all taxons investigated are usually not available in this field. Reference materials can be obtained from specific botanical, microbiological and zoological collections as well as from recognized scientific or research institutes working with those taxons or from other verified collections/samples (e.g. veterinary or zoological). Internal field specific collections provided by a forensic laboratory are accepted.
If a reference database contains non-domesticated species, voucher specimens should be used. The database should contain specimens that belong to the relevant population, both geographically and temporally.
There is no consensus regarding reference DNA sequences (e.g. mtDNA) for non-human species in the forensic community. The first sequence scientifically published or deposited in an online scientific database is typically accepted as a reference DNA sequence to which questioned sequences are compared (Budowle et al., 2005). This can be problematic for haplotyping if the reference sequence describes a rare haplotype or if the reference sequence is shorter or misses informative sites, making annotation sensitive to mistakes.

Internal and positive controls should be validated before being used in casework. Technical references are provided by the manufacturers.

4.4. Accommodation and environmental conditions

Incompatible activities within a laboratory (e.g. pre- and post-PCR procedures and samples; trace and reference material, between different pieces of evidence) should be separated in order to prevent contamination and carry-over.
Special care is recommended for the contamination caused by the environment and/or by reagents such as:

- bovine serum albumin in reagents
- pollen and spore contamination through the air
- bacterial DNA remnants in polymerases

4.5. Materials and Reagents

Changing the supplier of materials and reagents (e.g. buffers) may change the results significantly (e.g. overcome null alleles) and demands a review of the validation process.

5. METHODS

This chapter will not provide any standard operation procedures but provides guidance on systematic approaches to be followed in field specific examinations. It gives minimum requirements and recommendations for the selection of adequate analysis protocols and the order of their application in forensic casework. Typical pitfalls will be demonstrated.

5.1. General aspect

Three questions will be addressed regarding NHBT in forensic casework as well as methods and approaches to answer them:

(i) What is it?
(ii) To whom or to what does it belong?
(iii) Where does it come from?

The first question (i) can be answered by determining the family, genus, species (i.e. a taxon) name or a haplotype. Taxonomic typing can be used to find out whether a crime has been committed (e.g. the handling of illegal plants, mushrooms, wildlife) or, in a closed setting, to exclude or include its origin from the crime scene.
The second question (ii) requires techniques with greater discriminating power to individualize a trace or identify its origin. A match may indicate a link between a piece of evidence and an individual donor. The first stage involves taxonomic typing, while additional typing with other markers is required (e.g. STRs, SNPs or mtDNA analysis/haplotyping) in order to determine whether traces have the same donor.
The third question (iii) is to establish the geographical and/or genealogical origin of certain traces. This can be achieved by assigning traces to a certain population of origin, or by taxonomic typing if species are restricted to certain geographical areas.
In addition to these three main questions, more unusual questions can be asked. The evidence may be assigned to a group of individuals within a species, for example a cultivar, a subspecies, a chemotype or male and female individuals. The origin of a sample may be determined by identifying the composition of microbial taxa (microbiome) and assigning it to certain area of origin.

Several DNA-sources within the cell can be used: nuclear DNA (ncDNA), mitochondrial DNA (mtDNA) and chloroplast DNA (cpDNA), along with RNA and plasmids. Each of these sources has its own defining characteristics limitations, applicability and final evidential value. Differences in inheritance and availability in different tissues will influence the efficiency of the examination.

Non-human forensics – in contrast to human forensics – requires the development of new specific markers for every newly investigated group of organisms, as most markers are highly specific and work only in a limited set of related species.
The different reproductive strategies of organisms also play an important role in how the DNA/RNA methods can be applied to taxonomic typing, chemotyping, individualisation or geographical/genealogical assignment of NHBT.

5.2. Plan of examination

All examinations should be implemented following a plan of analyses and the results should finally provide adequate answers to the forensic questions identified. The plan must consider

the potential for destruction of evidential material during analysis and/or the potential need for further analysis. Where limited evidential material is available, the efficiency of the chosen techniques should be considered.

The methods applied may include (species-specific or taxon-specific) PCR, analysis of short tandem repeats (STRs), nucleotide-sequencing or hybridization techniques (addressing DNA polymorphisms e.g. SNPs and INDELs). If the respective technique has not been applied in forensic investigations previously a validation process is required (see Chapter 6).
Adequate presumptive testing is required to allow the minimum sampling and maximum efficiency of the molecular method. Before selecting appropriate techniques, all prior information must be carefully considered.
Positive and negative controls are required for each analytical technique and may help to detect inhibiting agents or the loss in activity of reagents. The implementation of extraction controls may be helpful to detect contamination during the stages of analysis. The selected analysis should be appropriate for the quantity of the target DNA/RNA in the sample and genetic divergence of the population.
Reference samples and forensic databases are of key importance, however the potential for misleading information must be considered if using open source databases. The databases applied for individualization as well as for taxonomic or population assignment should cover a broad spectrum of species, relevant populations and sufficient variation within populations, respectively. Determination of the origin also requires ecological expertise.

5.3. Demands on the analytical method

(i) What is it? Taxon identification

The first step in any of the suggested procedures is sample preparation. Possible subsequent analytical methods have to be envisaged and obvious characteristics of the sample (e.g. is it a mixture?) may have an impact on the preparation method (e.g. should it be homogenized or pooled?). The limitation of a forensic sample batch has to be taken into consideration for all subsequent analyses.
DNA (or RNA) is then extracted from the prepared sample and quantified. This may be assessed by various quantification techniques such as spectroscopy, fluorometry, gel electrophoresis or real-time PCR, each with advantages and/or disadvantages. However these procedures will consume a certain portion of the (potentially limited) sample and the measured values represent total DNA (RNA), not necessarily target DNA (RNA).

If the evidential sample does not provide any morphological indications regarding the kingdom or family of the donor and no prior information is available, the amplification of a region with taxon-specific polymorphism using universal primers and subsequent sequencing of the amplicon is the general approach.

The following target sequences are currently recommended:

- Cyt b, COI, 12S and 16S rRNA genes (mtDNA; for animals/mammals)
- ITS (ncDNA), trnL-trnF, rbCL/matK, psbA-trnH (cpDNA; for plants)
- ITS and LSU (ncDNA; for fungi)
- 16S rRNA genes (bacterial DNA)

The markers used by the members of the APST are listed in the ENFSI APST Working Group database, held on the APST site of the ENFSI intranet.

If the DNA sample is degraded and the above mentioned sequences cannot be amplified, non-coding D-Loop sequences (mtDNA) may be chosen as an adequate target for mammal species. Since universal or degenerate primers are commonly used in this approach, the target quantity (i.e. copy number of ncDNA and mtDNA) of the extracted sample may have an impact on the results (e.g. nuclear pseudogene copies of a mtDNA target may be amplified) or preferential amplification of one target in a mixture of several may occur and should be considered during evaluation.

If a readable sequence is achieved by this procedure a database must be searched to compare the sequence to similar or identical sequences derived from annotated sources. Both the National Institutes of Health (NIH) genetic sequence database GenBank (*http://www.ncbi.nlm.nih.gov/genbank/*) and the Barcoding of Life (BOLD) Systems databases (*http://www.boldsystems.org/*) provide public access. GenBank (NCBI) is cross linked to the DNA DataBank of Japan (DDBJ) and the European Molecular Biology Laboratory (EMBL) and harbours a wide variety of different sequences, but it lacks editorial surveillance and may contain incorrect annotations. The BOLD databases are annotated by taxonomists and reference samples are provided upon request. However these databases comprise only a limited set of genomic targets and sequence alignments are partly inaccessible. Next to these, various smaller, specialized databases exist which may also be used for forensic purposes (see also Chapter 12).
If verified reference samples of the species (taxon) indicated by a database match are available, additional sequencing of these is recommended. A direct sequence alignment may confirm (or precise) species (taxon) assignment (Linacre et al., 2011). Species assignment will be affirmed by the exclusion of the nearest neighbours.

If the evidence sample consists of a mixture of DNA templates, the interpretation of sequencing results is hampered. If the processing of raw data and manual sequence analysis (e.g. extraction of predominant signals from a mixture) does not provide readable sequences, the following alternative approaches for taxon identification may be applicable:

- Switching from nuclear to less ubiquitous non-nuclear targets that are characteristic of one part of the mixture (e.g. cpDNA sequences which are restricted to green plant tissues).

- The application of species (taxon) specific PCR for nuclear or non-nuclear (mtDNA or cpDNA) target sequences. This may be applied as a single analysis or in combined sets. For example, species (taxon) specific primers can be used in combination with conserved priming sites and in multiplex PCR reactions as available for mammals (Tobe and Linacre, 2008 and Pereira et al., 2010).

Taxon-specific PCR markers may also be chosen as a direct approach where specific species (taxons) are to be excluded or identified: more prior information or prominent morphological features narrow down the assumed set of taxa.

Specific nucleotide positions may be investigated for taxon-specific polymorphisms in a SNP assay. Next to sequencing, techniques based on primer elongation, such as minisequencing (SNaPshot analysis), allow the interrogation of single SNP positions.

Hybridization techniques may be applied for taxon identification. Microarrays with species (taxon) specific oligonucleotide probes (e.g. those derived from bacterial 16S rRNA gene sequences) fixed on solid supports are most commonly used. The oligonucleotide chip may be self-designed for a choice of targets and/or the oligonucleotide sequence that is presented on the surface of the slides. The oligonucleotides are hybridized to DNA or RNA that has been extracted from the forensic sample.

The investigation of regions revealing taxon-specific polymorphism may also use next generation sequencing (NGS) techniques. Mixed template sources can be resolved since these techniques display the nucleotide sequence of every single molecule that is subject to the sequencing reaction in parallel.

Should the forensic question require an individual identification next to taxon identification, an immediate STR analysis using taxon specific primers may answer both questions at the same time and economize the use of the sample.

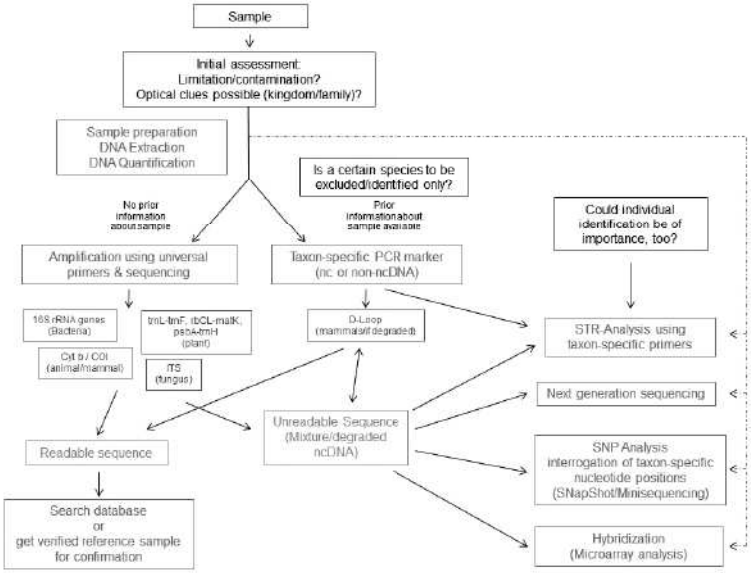

Fig.1: Taxon identification - What is it? Overview of general working stages and possible strategies. The flow diagram shows questions and information necessary for a decision (black), helpful techniques (red), suggested markers (blue) and data generated (green). The suggested flow of work is indicated by arrows, however this is not obligatory.

(ii) To whom or to what does it belong? Individualization

For an individual identification of a NHBT the genetic information gained from the sample has to be as distinctive as possible and should subsequently be compared to a reference sample for the assumed individual.

The most powerful methods for this task are taxon-specific STR markers (preferably tetrameric repeat systems) and/or polymorph nucleotide positions (SNPs) by PCR and capillary electrophoresis, Sanger Sequencing or NGS.

In some cases markers have previously been used in forensic applications (e.g. cattle, pigs, horses, dogs, cats, deer, birds of prey) and if not, existing marker-sets may be reformed and (re)-constructed (see the database of the ENFSI APST Working Group, held on the APST site of the ENFSI intranet). Those analyses which apply new markers that have not been used for forensic applications previously require preliminary validation in accordance with forensic standards, concerning both unambiguous allele identification and inheritance. The wide range of living forms with special genetic phenomena requires additional information or contribution from dedicated specialists.

The analyses result in a set of combined alleles of the chosen genetic markers that are recorded as genetic profiles. The necessary number of markers for individualization depends on the given species, populations and the number of alleles recorded. The obtained genetic profiles can be classified to full, partial or mixed profiles (with or without single allele drop in/ out; see Chapter 6). Those suitable for comparison lead to either a match (identical alleles in all investigated markers) or mismatch (divergence of alleles). The latter leads to an exclusion (i.e. the questioned and known samples are not derived from the same individual). It must be considered that several individuals may share the same genotype due to the prevalent mode of propagation (see (iii)).

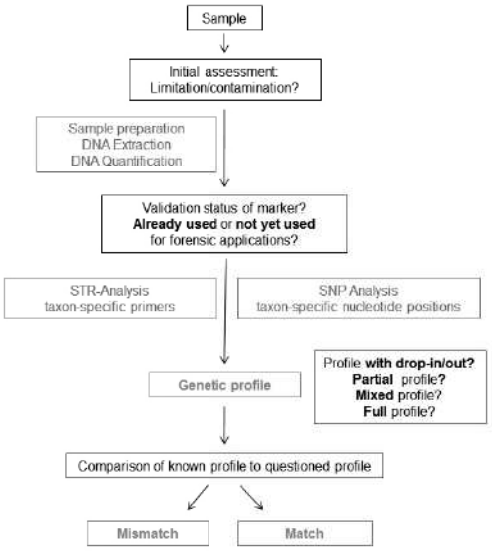

Fig.2: Individual identification - To whom or to what does it belong? Overview of general working stages and possible strategies. The flow diagram shows questions and information necessary for a decision (black), helpful techniques (red) and data generated (green). The suggested flow of work is indicated by arrows, however this is not obligatory.

(iii) Where does it come from? Geographical or genealogical origin

Determination of a population of origin or the geographical origin

For some forensic questions a certain “group of individuals” has to be determined from which the sample in question may originate. This could either be a population located in a specific geographical area or a distinct genealogical group such as a breeding lineage or direct kinship descent.

Two factors may have an impact on the strategy and the marker system chosen for the determination of the population of origin.

a) The habitat

- If the sample in question belongs to a species that is restricted to a fixed habitat, species identification (as shown in i) is a sufficient strategy to determine the geographical origin of a representative of this species.
- If a species is more prevalent geographically or if the habitat harbours different populations, the population of origin must be assessed in order to identify the geographical origin of the sample. To assign a sample to a certain population a DNA marker system is required which detects genetically distinct alleles shared by all individuals of the target population and is absent or less abundant in other candidate populations and thereby differentiates between both; however such a marker system may not always be available.

b) Mode of reproduction

Different levels of genetic variation in the progeny result from various modes of reproduction in non-human organisms, especially in plants, or from breeding of domesticated species. This may lead to populations of varying divergence:

- Cross-breeding (outbreeding) species mostly reveal large genetic differences in their populations and therefore a distinct genotype allows individual identification (see ii).
- In selfing species, inbreeding populations and species performing clonal propagation or apomixes, the genetic variation within the offspring is reduced and individuals with identical genotypes may occur.

The degree of genetic exchange between different populations plays an important role when determing the strategy used for identification, therefore the existence of population databases and allele frequencies reflecting the diversity among and between individuals is of major importance. However the quality and quantity of population data can be diverse. In particular, the collection of population data for rare or endangered species may be problematic.

- If the existing population data reveal a highly divergent population that is isolated from others (e.g. by geographical barriers), without gene exchange with other populations, members will share the same type of genetic differences and exhibit discrete types of variation when compared to other populations. Fixed alleles exist and may be detected by SNP or sequence variation analysis of mtDNA. Hypervariable D-loop haplotypes in particular are found to be specific for certain populations, therefore all other populations can be excluded as the source of origin for a matching sample.
- If the existing population data reveal a less differentiated population with some limited gene exchange, mtDNA variation might be rare and markers located in the nuclear genome, such as STRs and SNPs have to be employed. Some alleles may be distributed across populations, however different frequencies between populations may allow differentiation.
- If no population data is available or the population reveals a high gene exchange, (i.e. the allele frequencies are more or less equal among different populations) no genetic method can be applied and other methods (e.g. isotope analysis) have to be employed to determine the geographical origin of the sample.

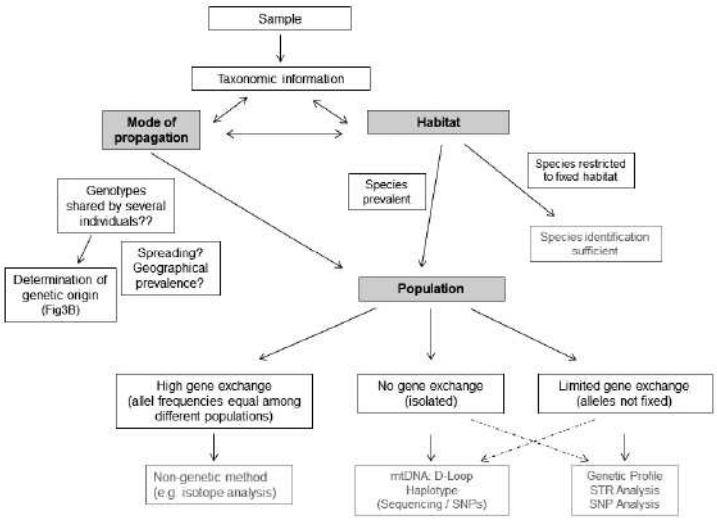

Fig.3A: Determination of the population of origin (geographical origin) - Where does it come from? Overview of general working stages and possible strategies. The flow diagram shows questions and information necessary for a decision (black) and helpful techniques (red). Important information is shaded grey. The suggested flow of work is indicated by arrows, however this is not obligatory.

Determination of a genetic origin (kinship or breeding lineage analysis)
The ability to verify or refute familial relatedness is an important application for non-human forensic investigations. In wildlife investigations such as illegal trading or poaching, it may be necessary to differentiate between animals bred in captivity and wild animals. This may also apply for plant species (e.g. vegetative propagation by cuttings). A common genetic origin or parent-offspring relationship can be evidenced by genetic markers that are inherited sexually from one generation to the next or passed on to a non-sexual descendant.

In case of sexual reproduction in outbreeding diploid species, parentage and other kinship relations can easily be confirmed or excluded using SNP or STR markers. The alleles in the DNA-profile of the offspring individual(s) must be present in its putative parents with each parent contributing one allele per marker. If alleles occur which do not correspond to the putative parental profiles, they can be excluded. More complex kinship analysis, for example sibling identification, may be achieved with analogous approaches.
For the exclusion of kinship relations, no population data is required. However, the variability of genetic markers is based on heritable (rare) mutation events (see Chapter 12).

The existence of other polyploidy levels demands adaptations to the approach described. Special attention has to be paid to different forms of propagation (see above). If identical (or very similar) genotypes are shared by several individuals, a common genetic origin may be deduced directly. However the exact genesis of the descendants (e.g. offspring being produced by clonal propagation, repeated selfing or apomixes) cannot be reconstructed without further knowledge about the species.

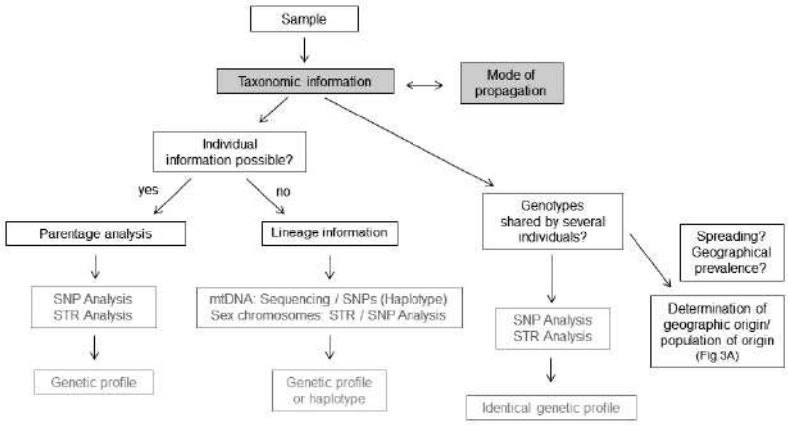

Fig.3B: Determination of the genetic origin (kinship analysis or breeding lineage) - Where does it come from? Overview of general working stages and possible strategies. The flow diagram shows questions and information necessary for a decision (black), helpful techniques (red) and data generated (green). Important information is shaded grey. The suggested flow of work is indicated by arrows, however this is not obligatory.

In the absence of one or both parents, or if descendants have to be retraced over several generations, the analysis of a breeding lineage may be more convenient than an STR or SNP approach. Lineage analysis may be conducted by sequence analysis or SNP analysis of the entire mtDNA or its hypervariable (control) region (D-Loop) in animals. Due to its uniparental (mostly maternal) heritage, the resulting haplotype should be matrilineal passed over generations unaltered or only slightly altered. The higher robustness and amount of mtDNA compared to nuclear DNA can be advantageous for forensic applications because it may enable the investigation of less well-preserved tissues (e.g. degraded bone material) if no other material is available.

Markers residing in the non-recombining regions of sex chromosomes, such as mammalian Y-chromosomes, enable ancestral tracking through males. Z-chromosomes in birds and amphibians are inherited uniparentally and may also be used for lineage analysis in some species. For these regions of the nuclear genome, STR or SNP analysis is applied and leads to a haploid profile. Due to duplications and translocations more than one allele may be observed for some marker loci. These markers are located in non-recombining regions and are consequently inherited as a complete haplotype. This should be considered when estimating the statistical significance of these markers (see Chapter 12).

Determination of other groups of individuals within a species

The identification of certain groups of individuals within a species may also address functional properties, for example plant cultivars or animal subspecies revealing a certain phenotype or producing certain active components (e.g. pathogenic microbial strains or the THC/CBD chemotype in *Cannabis*). Generally, group-specific mutations or genes in the nuclear genome are detected by selective PCR or sequencing the characteristic region.
A single site or multiple sites (e.g. a gene responsible for a monogenic trait, a specific transgene, a group of toxin genes) harbouring mutations can be targeted. Genetic variation in

these loci should be linked either genetically to the trait of interest (pathogenicity, chemotype) or their functionality should have been authenticated independently. Group-specific identification also comprises sex determination in dioecious, sexually propagating species. Male and female individuals can be determined by targeting polymorphic markers in sex determining regions or closely linked to such.

Complex traits are most often multigenic and markers linked to the phenotype allow only statistical predictions. In this case a reliable statistical fundament is necessary for a forensic application.

Determination of the composition of a population of microorganisms: Microbiome analysis

The composition of taxa represented in a community of microorganisms allows a sample to be assigned to an environmental source if it is stable and distinctive. Due to its large microbial diversity and the influence of various biotic and abiotic factors, provenance-dependent microbial DNA profiles can be used to determine the origin of non-human cell material found on the human body, for example on the skin or in the digestive system, or the geographic origins of soil traces.

Direct DNA-extraction and subsequent PCR-based techniques from these bacterial populations are available, therefore the microbial diversity can be visualized without the loss of non-culturable species. The visualization of abundant and less abundant species within the community makes this method very sensitive to cut off values and diluting effects.

The method of extraction is important, as some bacteria (e.g. Gram positive bacteria) and spores are hard to lyse, or contained within a rigid matrix such as instant soil. Methods with stringent lysing steps, for example bead-beating, are very suitable for extracting DNA from cells which are difficult to lyse or those within rigid matrices. The choice of extraction method will influence the final DNA-profile of the microbial population.

For most downstream analyses the amount of input DNA needs to be quantified as different matrices can result in a broad range of bacterial DNA-concentrations and will subsequently influence the final DNA-profile. Certain soil or faecal samples can contain large amounts of bacterial DNA.

The 16S rRNA genes (bacteria) or 18S/23S rRNA genes (fungi) can be used as target loci for community profiling. For more differentiation within faecal *Enterobacteriaceae* the GroEL gene can be used.

The approach may also involve plant and metazoa targets to investigate a broader spectrum of living beings (Giampaoli et al., 2014).

Following PCR, methods to visualize the population typically include digestive techniques, for example terminal restriction fragment length polymorphism (t-RFLP, RISA), hybridization techniques such as microarray analysis or massive parallel sequencing assays (NGS). All techniques have different resolutions and it is important to choose the technique(s) with the desired resolution.

- With t-RFLP/RISA no taxonomic information on the microorganisms present will be generated. Nonetheless, these techniques are suitable to compare samples in order to determine a possible common source. PCR can be performed with one or two labelled primers (depending on the regions amplified one labelled primer may be sufficient). The choice of primer combination and restriction enzyme(s) determines how many restriction fragments will be generated.

- With microarray analysis, taxonomy information becomes available. For some applications commercial microarrays are available, but custom made arrays may also be suitable. 16S rDNA probes can be validated against the Ribosomal Database Project (RDP) database while probes targeting other genes or organisms can be validated using public or private databases.

- Next generation sequencing can give insight into a population at a very high resolution. Amplicon sequencing of the 16S rRNA genes can be used for taxonomic determination to at least genus level, while shotgun sequencing may lead to lower level taxonomic determination. The desired length of the reads depends on the application for which NGS is used; for species identification longer reads can provide more information. Extremely high resolution NGS may not always be necessary for forensic investigations.

5.4. Peer Review

The peer reviewing should be organized by the laboratory quality system. If possible, all critical processes such as the analytical plan of methods applied and the interpretation of data should undergo a peer review by a second practitioner.

All peer review must be documented in the case record.

6. VALIDATION AND ESTIMATION OF UNCERTAINTY OF MEASUREMENT

6.1. Validation

Techniques applied in forensic investigations require preliminary validation according to forensic standards, such as ISFG recommendations (Linacre et al., 2011). The method of validation depends on the specific technique and its application. The application of novel markers requires preliminary clarification or determination of their genetic background. The specific genetics of the organism investigated may require different forms of validation. Generally validation can be divided into biological validation (of species markers, STR-systems or SNPs, including their interpretation), technical validation of analytical methods using dedicated equipment and the validation of interpretation. These aspects have been previously documented for the field of non-human DNA-typing by Budowle et al. 2005, Linacre et al. 2011, and the SWGWILD Standards and Guidelines 2012.

6.1.1. Biological validation

Regarding the validation of DNA-markers for certain purposes (taxonomic identification, within species “individualization”, parental analysis), known biological parameters should be evaluated:

- Genetic source of the DNA-marker

The genetic source of a DNA-marker will influence the inheritance, as will be discussed in the next paragraph. However, if the genetic source is a multiple copy locus, length heteroplasmy and nucleotide heteroplasmy can occur. In addition, copies of mtDNA sequences may be present in the nuclear genome. These features may be encountered during validation and can vary between individuals, species and populations, without affecting the applicability of the DNA-marker. However, it should be documented if these types of variance should be expected and the resulting consequences.

- Type of inheritance and propagation (Mendelian, maternal/paternal, clonal, self-pollination, etc.)

The type of inheritance affects the evidential value of the outcomes of the analysis. During validation of the marker, it should be known if inheritance is independent from other markers (loci), and whether it follows Mendelian inheritance or not, due to its location on chloroplasts, mitochondrion or sex-chromosome.
In plants, reproduction by self-pollination, cloning and the inheritance of chloroplasts (maternal in angiosperms, paternal in gymnosperms) can influence the appearance of the marker/allele in the next generation or in the population.

- Biological variation (within a single individual, between individuals within species, between species, between populations)

In order to assess the applicability of a marker for species identification, individualization or parentage testing, it is important to document the behaviour of the marker within the relevant population. Previously published data about the biological variation of the marker should be verified in the population of interest from a forensic point of view.

Markers for species identification should be tested for both the intra-species variation and the inter-species variation (to specify differences with the closest neighbours).
For markers for individualization (STR), the number of alleles, allele frequencies in the relevant population, etc. has to be determined. This includes identifying whether genotype-frequencies are in Hardy-Weinberg-equilibrium, giving insights in factors such as inbreeding, bottle-necks, ecological barriers and the presence of null-alleles.
If STR markers are used for parental testing, attention should be paid to inheritance and mutation rates. In addition, the structure of a repeat and the ability to unambiguously identify alleles should be evaluated.
Repeats of less than three nucleotides, although extensively described in literature, should be avoided since they produce stutter artefacts which are problematic when applied in a forensic setting (e.g. poorly conserved samples, limited trace material).

For biological validation of DNA markers, recommendations published by Linacre et al. 2011 are applicable for animal DNA testing. This includes selection of the appropriate marker, primer-sequences, intra- and interspecies studies for taxonomic typing and published primer-specificity and reproducibility, mutation probabilities for STR-alleles, allele frequencies and kinship factors for identity testing.
When animal DNA-testing is conducted routinely and/or inter-lab-comparisons are required, recommendations regarding the use of tetrameric repeats and allelic ladders can be of use.

Similar criteria are in place for plant typing, although special strategies for reproduction (clonal, self-pollination) must be taken into account.

6.1.2. Technical validation

The technical validation refers to the optimization and standardization of an analytical technique with the chosen reagents, equipment and controls. For technical validation, many criteria and recommendations can be found, for example in the ENFSI-QCC document “Guidelines for the single laboratory Validation of Instrumental and Human Based Methods in Forensic Science” (2014). It should be recognized that some criteria and recommendations are written for quantitative methods. Since DNA-analysis can also be only qualitative, especially when end-point PCR is involved, as for most of the STR-analysis, parental testing and taxon identification, some of the validation parameters particular to quantitative methods are not applicable.

For technical validation, additional field-specific recommendations have been previously documented (Budowle et al. 2005, Linacre et al. 2011, and the SWGWILD Standards and Guidelines).

6.1.3. Validation of interpretation

For species identification, the evaluation of specific markers, base-calling and/or the comparison of sequences should be validated.

For STR-analysis, the identification of alleles, comparison of profiles, assessment of null-alleles, allelic drop-in, allelic drop-out and the production of artefacts should be validated (Gill et al., 1997). For the interpretation of mtDNA typing guidelines have been reported and should be validated for the application to NHBT (Bär et al., 2000).

Mutation events may have occurred in the generation selected for analysis, which may in turn cause inconsistencies at a limited number of loci while all other loci show genetic consistency: therefore not every false allele indicates false parentage but may be due to a (rare) mutation event.
Mutation rates are marker and locus specific and should be assessed by screening pedigree studies: these should then be incorporated into parentage analysis. However this is not often possible for wildlife or rare species, therefore profile interpretation must be undertaken with caution. If mutation rates are not available, a conservative estimate of 10^{-3} mutations per generation has been suggested for STR markers and differences at more than two loci per 10-12 STR markers as a threshold for parental exclusion (Dawnay et al., 2008).
In general it can be stated that when analyses are applied in only a single case or on rarely occurring traces, allelic ladders and standardized nomenclature are not yet required. If results have to be compared with other laboratories or earlier results in databases, these tools are strongly recommended.

6.2. Estimation of uncertainty of measurement

When applying quantitative PCR, the standard deviation of the CT-values should be determined, translated to actual deviation in DNA-concentration and reported. For qualitative methods the estimation of uncertainty of measurement is an integral part of the validation of interpretation process (see 6.1.3). Possible sources of uncertainty arise during sample preparation, DNA extraction, purification, amplification, detection and interpretation.

7. PROFICIENCY TESTING

Proficiency tests should be used to test and assure the quality of the molecular genetic analyses of NHBT in forensics. A list of currently available PT/CE schemes as put together by the QCC is available at the ENFSI Secretariat. “Guidance on the conduct of proficiency tests and collaborative exercises within ENFSI” provides information for the ENFSI Expert Working Groups (EWGs) on how to organise effective proficiency tests (PTs) and collaborative exercises (CEs) for their members.

PT samples could be part of validation experiments and/or harmonization processes between different laboratories applying the same, or different, analytical methods to produce results for a particular forensic examination.

If there are no proficiency tests available for a specific field, the user of the particular analytical

method in question must conduct a blind trial using reference samples to ensure the reliability of the results obtained as part of an internal quality assessment.

Within existing methods, these tests are used to identify potential sources of error, produce corrective measures and optimize the potential for improvement of any quality management system. The quality measures are designed to demonstrate the accuracy and precision of results and the limits of detection of the analytical method tested.

Proficiency testing serves not only to evaluate the chosen method, but also to assess the performance of the laboratory and their ability to reach appropriate conclusions following the application of the analytical method in question. (see also "The GEDNAP proficiency testing system, 2013).

For the examination of samples the laboratories are expected to follow the international guidelines for forensic DNA analyses and to include all necessary controls. For DNA extraction and amplification, adequate controls (e.g. positive, negative and reagent controls) must be processed in parallel. If there are no existing international guidelines for newly adapted methodologies the rules must be defined by the coordinator of the PT/CE or an alternative advisory group.

Laboratories are asked to retain an adequate quantity of the test sample for second opinion testing should there be any disagreement over the identity of the sample or concerns regarding contamination prior to the sample being received by the analyst/investigator.

8. HANDLING ITEMS

The collection, packaging and handling of samples from animals, plants, fungi and soil for molecular analysis requires the same standards as the investigation of any incident. The integrity of the item and the traceability of evidence (i.e. chain of custody) are required to ensure that the developing case is not challenged as a result of improper handling procedures.
It is important to note that this manual is not a guide for crime scene investigations nor for crime scene investigators. General information is available at the Forensic Science Network Scene of Crime Working Group.

8.1. At the scene

Crime scene investigation is a systematic process that aims to record the scene as it is first encountered, recognize and collect all physical evidence which may be relevant to the solution of the case. Preserving the chain of custody is of vital importance to this process. This refers to the careful and chronological documentation of the handling and analysis of evidentiary materials during a case. Throughout the process, it is crucial to be able to demonstrate every single step undertaken to ensure the traceability and continuity of the evidence from the crime scene to the courtroom.
When handling evidence at any stage of an investigation, possible contamination of samples must be avoided. Gloves should be changed when handling different objects to prevent cross contamination between samples. Crime scene investigators have to be particularly aware of non-human biological contamination which can, for example, be introduced by police dogs or pollen in the air.
Personnel working at crime scenes and handling samples may be exposed to various health and safety hazards. These may include chemicals, biological material (e.g. blood and body

fluids), explosives and other risks. Therefore health and safety procedures should remain a priority throughout the process.

8.1.1. Recognition of evidence

For the detection of biological samples (blood, saliva, hair, tissue, bones, plant material, etc.) on tools, clothes, etc. during the crime scene investigation, visual inspection will often be adequate. In some cases the use of special lighting may be required.
Before collecting the samples the presence of other types of forensic evidence, for example, human DNA and fingerprints, must be taken in consideration: care must be taken in preserving these.

8.1.2. Collection of evidence

Whenever possible, stains should not be removed from an object at the crime scene; the object should instead be collected. If the stain is on an immovable object, it should be cut out and removed if possible (e.g. a section of carpet) or collected with a cotton swab manufactured for forensic purposes (e.g. from a solid or fixed floor surface). In these cases, control samples should also be collected to allow the evaluation of the genetic background.

8.1.3. Preservation and packaging

All NHBT should be stored in a cold, dark and dry environment. Living material should be treated according to their specific needs. Moistened stains should be dried as soon as possible or frozen immediately.
Liquid body fluids (blood, saliva, semen, urine) must be stored in specific containers and transported at cool environment. Animal tissues, bones and teeth should be cooled. Plant material may be dried. For longer storage soil samples should be frozen (-20°C).

8.1.4. Transport

The evidence should be transported to the laboratory as soon as possible. Repeated thawing and freezing cycles should be avoided.

8.2. In the laboratory

Identifying information on articles of evidence must be checked on receipt to ensure it is consistent with the submission form.
Stain materials should be collected, identified and analyzed using appropriate DNA-techniques. DNA-analysis techniques are highly sensitive and may be affected by contamination, therefore contamination and carry-over should be avoided throughout the laboratory investigation. In particular, contamination of trace material with reference samples must be avoided by handling these items separately in time and/or space.

Where possible, leave a sufficient quantity of raw sample or extract for a second test or subsequent investigations.

DNA extracts should be stored between -20°C and -80°C for long term storage or at 4°C for a short period of time. Repeated thawing and freezing cycles should be avoided.

9. INITIAL ASSESSMENT

Every condition and action (e.g. interruption of the cooling chain, decomposition, degradation, fermentation, chemical treatment) which may affect the quality and quantity of DNA/RNA within

the trace should be taken into consideration when selecting methods to be used for the analysis and interpretation of results.

10. PRIORITISATION AND SEQUENCE OF EXAMINATIONS

Guidance for establishing priorities and sequences of examinations of NHBT in the laboratory setting are provided in Chapter 5 (Figure 1 – 4).
If morphological or chemical investigation is to be done in addition to PCR based examinations, the risk of contamination must be considered and minimised. Especially chemical pre-treatments/tests or UV-light may cause degradation of target DNA/RNA.

In general, NHBT exhibits are only analysed using molecular methods, not as part of a multidisciplinary investigation. If other bodies of professionals are to analyse the trace evidence, the splitting of samples should be considered in advance.
If the carrier of the traces itself is undergoing multidisciplinary investigation, all relevant experts should determine the most appropriate sequence of examinations as dictated by each individual case.

11. RECONSTRUCTION OF EVENTS

Not applicable

12. EVALUATION AND INTERPRETATION

Recommendations, standards and guidelines on evaluation and interpretation have been described previously for the field of non-human DNA-typing by Budowle et al. 2005, Linacre et al. 2011 and the SWGWILD Standards and Guidelines 2012.

12.1. Evaluation of data

The method of evaluation must be appropriate for the applied analysis. The results have to be derived unambiguously from sequence or/and length matches or differences. The evaluated result must be invariable to analytical conditions.

The analytical data should be evaluated before interpretation. Depending on the analysis performed, the following criteria may be considered:

- Automated base or allele calling should be checked manually and independently by two practitioners where possible. The manual adaptations/mutations should be recorded and stored together with the raw data.
- For sequencing, each position should be base called by at least two reactions, starting from different primers, if possible or applicable.
- For STR analysis, designation of the sample allele is only possible if the largest and smallest alleles for that sample fall within the range covered by the internal size standard.
- The positive and negative controls that were taken through the analytical process should be evaluated. If a control shows deviating results the practitioner should reject the results or consider this during the evaluation/interpretation of the data obtained for all other samples in the same batch. The lab should provide rules for how to handle control results.
- The risk that the obtained analytical result is caused by contamination should be evaluated. The contamination risk is sample type dependent:

  - Tissue samples that can be decontaminated (e.g. bone) vs liquid (e.g. blood spot)
  - Abundant tissue sample vs. trace samples with low DNA concentration
  - Mixtures vs. pure material
- Each laboratory should implement minimum thresholds for the acceptance of data. Examples of data quality indicators include quality scores, signal intensities, peak heights, and background levels.
- When interpreting mixtures, the number of contributors, the genetic background, contamination issues (e.g. the admixture of human DNA) and the technique chosen should be taken into consideration and alternative strategies or methods can be applied if necessary (e.g. the ploidy in plants will affect STR analysis, while the primers chosen may influence the ratio of the signals in Sanger-sequencing and multiplex reactions).
- Be aware of the presence of heteroplasmy in multiple copy loci (e.g. mtDNA). Criteria for the detection of heteroplasmy should be stated (e.g. noise to signal ratio). The tissue type of the analysed sample and the heteroplasmy rate in the studied sequence has to be considered: always be conservative in the data interpretation as heteroplasmy may be present in the sample but below the detection limit of the used technique.
- In case of mtDNA analysis, the (co-)amplification of a nuclear pseudogene must be considered.

12.2. Interpretation of data

The interpretation of data obtained from non-human biological samples in forensic casework is discussed with reference to the three main forensic questions, as stated in Chapter 5: (i) what is it? (ii) to whom or what does it belong to? and (iii) where does it come from?

Multiple factors can make the interpretation of non-human DNA data challenging: a wide variety of species could be encountered during case work, multiple sources of DNA can be analysed and many different analytical techniques could be used.

Some species/populations have not been extensively studied and are not often investigated in forensic cases. The interpretation of DNA data from such specimens cannot rely on reference data to gain insight into phenomena such as genetic sub-structuring of a population, mutation rate of genetic markers, etc. The lack of genetic reference data therefore has an influence on the interpretation of the data: conclusions should be drawn on a case by case basis, and carefully detailed in the report.

During interpretation, the use of DNA data from reference samples or forensic databases is recommended if available. Special attention should be given to the pre-evaluation of open source databases, since biased or false data can negatively influence interpretation. Taxonomic classification, individualization and population assignment require dedicated databases which describe the relevant population and the variation within this population. Additional specific database requirements will be discussed for each application.

The availability of frequency data enables the statistical interpretation of results using likelihood ratios (LR) following Bayes Theorem. The frequency data must be supported by a representative reference population. In the absence of relevant data the nominal statistics cannot be provided and an estimated category of probability should instead be considered.

The interpretation of DNA data from biological traces in a forensic investigation may also require

ecological expertise regarding the particular species under investigation. The interpretation may therefore be a combination of both genetic and non-genetic data. The interaction between these two types of data is not within the scope of this BPM.

(i) What is it? Taxon identification

A sample will be assigned to a certain taxonomic level, while excluding all other species, genera and families. The taxonomic level achieved for the identification should be stated.

It is important to note that the concept of a species is not clearly defined.

Taxonomic classification based on DNA sequences compares the DNA sequences obtained from the samples with reference DNA sequences.

Locus/loci
The choice of locus/loci is level and taxon dependent and should be evaluated for each case.
If the practitioner has no experience with the questioned taxon, they should consult scientific literature to study the taxonomy or contact an external expert in the relevant scientific field to determine the most appropriate locus/loci.
The locus/loci should enable the identification of the unknown taxon among those that are close genetic relatives. To achieve this, the evolutionary relationships of the taxons in question should be considered (e.g. a faster evolving gene would be recommended for groups that diverged relatively recently from a common ancestor).
The primers selected may be universal or more specific, depending on the taxonomic level to be determined.
The use of multiple loci should be considered, particularly for those taxons where a single locus is insufficient for classification (e.g. plants). The loci should be evaluated in combination taking into account their individual mutation-rate.
The markers used by the members of the APST are listed in the ENFSI APST Working Group database, held on the APST site of the ENFSI intranet.

Using reference databases
The database selected should contain the taxons under consideration along with species known to be closely related genetically. If not, the limitations of the reference data should be described.
The database should also contain multiple specimens from each of the taxons.
Determining the origin of a trace requires not only relevant population data but also ecological expertise. If necessary, an expert in that particular field should be contacted.

The obtained DNA sequences should be compared to an appropriate reference database by sequence match.

The practitioner should check the intra-species variation for this sequence and the distribution of genetic distances among closest relatives.

The parameters used for a search in a public database should be evaluated per taxon and modified if needed.

As the species (taxon) of origin may not be present in the database, non-matches should be interpreted carefully for species (taxon) attribution. Identification to the taxonomic level can be

evaluated by the construction of a phylogenetic tree that contains the unknown sample and reference sequences from the most appropriate species available. The branches should include a measure of statistical support that is known to be taxonomically robust for the species in question.

The following phenomena may limit taxon delineation and should be considered during interpretation: hybridisation, introgression, homoplasy, cryptic species and nuclear integrants of mitochondrial and plastids DNA sequences (NUMTs). The appearance of pseudogenes should also be considered.

(ii) To whom or to what does it belong? Individualization

The goal of individualization is to match the questioned sample to a known sample using a probabilistic approach.

STR-loci
The STR-loci are tested for genetic linkage, polymorphic content and inheritance in a standard Mendelian manner. To reduce the presence of stutters, tetra-nucleotide repeats are preferred over di-nucleotide repeats. Stutter and heterozygous balance should be recorded.

STR-Allelic ladder
If a complete and comprehensive allelic ladder is available, the observed STR-alleles could be called by repeat numbers and should be recorded as such. This allows comparisons between different analyses, instruments and protocols.
If no allelic ladder is available, the alleles should be recorded in base pairs. The harmonization of the allele calls should be done with reference to known samples.

If a comparison of STR results between laboratories or another analytical system is necessary, a reference sample or an allelic ladder is required as a reference.

SNP
Like STR-loci the SNP's are tested for genetic linkage, polymorphic content and inheritance in a standard Mendelian manner. An allelic ladder is not needed.
The signal intensities of the currently used SNaPshot technique are less correlated to the amount of template DNA than for the STR-technique, which should be considered in the interpretation of sample mixtures.

Population database
A population database is used to estimate the allele frequency in the population which is appropriate for the specimen in question, if available. Rare alleles that are not found in the database should be considered during calculation of match probabilities (Johnson et al., 2014).
A population study would not be feasible where the specimen in question belongs to an endangered and protected species.
It is recommended that a database with sufficient variability is used: it should contain specimens that are expected not to be directly genetically related. If not, more loci or specimens should be added. The number of individuals in the consulted database should be indicated.

Interpretation for individualisation
The minimum number of DNA donors in the sample should be evaluated.

Consideration should be given to the observed versus the expected heterozygosity for each marker.
The possibility of stutter peaks, -A peaks, pull-up peaks, spikes, primer mismatches, and allelic drop out or drop in should also be considered.

Profiles of the questioned specimen should be compared with the known specimen.

In case of a profile match, the statistical support of this match should be calculated using the appropriate allelic database. This requires relevant data about markers and populations (frequency, distribution, variance, substructure, relatedness and linkage).

If required, the likelihood statistics can be expressed using Bayes` Theorem. The calculation of LR can be corrected using different factors (e.g. a kinship factor or theta value for inbreeding effects) if co-ancestry and/or population sub-structuring are present, if appropriate population data is available (Johnson et al., 2014). For domestic animals, breeding is controlled or influenced by humans, which ultimately results in inbreeding. Wildlife animals have 'at random' breeding, unless they are rare or geographically isolated, although the genetic variability of wildlife could also be (partially) human influenced (e.g. re-introduction of species in a habitat). It should be noted that there is not always a strict separation between domestic and wild animals (e.g. introgression of wild and domestic cats, dogs and pigs).

All limitations (e.g. unusual inheritance, mating behaviour, representativeness) of suboptimal relevant genetic data must be recorded transparently.

Mismatch with more than one allele must generally be evaluated as exclusion (i.e. the sample and reference are not from same individual).

(iii) Where does it come from? – Geographical or genealogical origin

Geographical origin: the association of a particular sample with a specific population from a known location, while excluding it from other populations

The assignment of a sample to a particular population can be based on species identification or on the variation of the selected marker allele frequencies among populations.

The DNA sequence or STR-profile originating from the questioned sample should be compared to data sets of known populations.

mtDNA
The Genetic distance between sample and population reference should be calculated from the degree of sequence similarities and is often displayed as a tree (for mtDNA coding region) or network (for mtDNA control region).

nuclear DNA
Sample assignment to a population by nuclear DNA markers could be based on the presence/ absence of fixed alleles or on the allele frequencies.
The candidate populations should be defined prior to analysis, and if not, the presence of population substructures in the reference data should be evaluated by a clustering test.
The consulted database should be representative for the considered geographic areas/

populations and should contain sufficient data, depending on the marker type used (SNP, STR) and its location (e.g. sex chromosome).

If the candidate populations can be distinguished by particular sequences or alleles the sample profile should be assessed on the presence or absence of these sequences/alleles.

If the population assignment relies on the allele frequencies, the probability that a sample profile belongs to a particular population should be calculated for each candidate population. The obtained probabilities should then be compared to each other.

If the populations of the particular species are genetically well characterized, different scenarios could be considered:
(1) If each individual reveals a different genotype, they can be identified and the population and geographical origin become known automatically.
(2) If a certain genotype is shared by various individuals their mobility and/or mode of dispersal (e.g. by roots, wind, animals, human breeders or trade activities) and their resulting population specificity and geographical prevalence must be taken into account.
(3) If individuals sharing the same genotype occupy large areas (e.g. plant species) - the geographical localization of a matching sample may not be well defined. For example, clonal *Cannabis sativa* (hemp) cuttings may be propagated on an industrial scale and distributed to various locations across Europe. Where individuals with identical genotypes are discovered in different geographical locations, this does not necessarily indicate a direct connection between the different plantations, this could instead be a common genetic origin. However, trade activities and routes may be retraced by such discoveries.
(4) If identical or closely related genotypes are shared by individuals from a limited local area (for example, in inbreeding plant species), or where occasional crossbreeding events introduce new genotypes in a certain location, a sufficient degree of genetic differentiation may be achieved to determine the population of origin (and probably also the geographical localization) (Koopmann et al., 2012).

For population assignment or the determination of the geographical origin of a specimen, the biogeographical data of the species in question should be considered. This includes natural habitats, territories, movement of males and females, and the human influences on the species, such as migration (instigated by human activities), breeding programs and re-introduction programs.

Determination of a genealogical origin (kinship or breeding lineage analysis)

Kinship analysis (Paternity testing by SNP or STR markers):
When dealing with species propagating by sexual reproduction, the profile of the questioned specimen should be compared with the pedigree.

When a match occurs, the statistical support can be calculated if an appropriate allelic database is available.

To exclude a kinship relationship, no population data is required. However, not every false allele indicates a false parentage as this may be due to a (rare) mutation event. Mutation rates are marker and locus specific, and ideally, should be assessed by screening pedigree databases

and then incorporated results into parentage analyses. However this is rarely possible for wildlife or rare species, therefore profile interpretation requires cautious interpretation. If mutation rates are not available, a conservative estimate of 10^{-3} mutations per generation has been suggested for STR markers (see Chapter 6.1.3).

Breeding lineage analysis:
Genetic markers located in non-recombining regions, which are consequently inherited as a complete haplotype, should be compared between the questioned sample and known specimens. This should be considered when interpreting the data.

To confirm a match, a population database is required. The random match probability of the questioned sample profile in the considered population can be estimated by verifying the frequency of this haplotype in an appropriate reference database.
The size of the database influences the final result and should therefore be mentioned.
A database for frequency estimations in a population requires a particular method of sampling, preferably at random and covering the whole considered population. Reference databases collected for studying genetic variety in a particular population do not always meet these criteria, and should hence be utilised with caution.

A non-match results in an exclusion. However, the phenomenon of heteroplasmy in mtDNA and the site specific variation rates of the haplotype markers should always be considered in the interpretation.

Determination of other groups of individuals within a species

The appearance of the characteristic marker signal or sequence allows the affiliation of the sample to a particular group of origin and excludes it from all other groups. However if a single locus is investigated, mutations in this locus must be taken into consideration (e.g. for sex markers). Group affiliation can be confirmed by complementing or additional markers/loci.
The existence of man-made modifications (e.g. by transgenes, knock-outs or breeding effects) should be considered, depending on the species investigated. Also naturally occurring phenomena like dioecious plants or hermaphrodite flowers have to be considered in sex determination.

Further knowledge about the background and peculiarities of the species may be necessary.

Microbiome analysis - Determination of the composition of a population of microorganisms

The final construction of a bacterial profile is very sensitive to the extraction method, applied PCR and subsequent applied software (including cut-off values and normalization) during data collection.
When a profile is generated, it may consist of data points which correspond with known specimens (analysed by NGS, microarrays) or with unknown specimens (analysed by tRFLP, RISA). It may in some cases be possible to assign a population to a certain origin (geographical or niche) using information at genus or species level without further processing. In many cases however, comparisons should be made with reference samples (databases). Generally this can only be achieved when the same methods have been applied to the questioned samples and reference samples. To make comparisons, an objective measure for the similarity between the bacterial profiles is required as an exact match will never be achieved due to biological and

technical variation and sampling effects. For this purpose, ecological distance measures (e.g. Bray Curtis) or other correlation coefficients can be used to express the similarity between samples (Quaack and Kuiper, 2011). Choosing an appropriate method depends on the dataset and the chosen data processing tools. This should be accompanied with a decision model to define which similarities must be present in samples to be considered to have the same origin (match) or not (mismatch), or when data is inconclusive. This can typically be achieved by building a database with known samples and plotting the acquired coefficients between samples from the same origin together with acquired coefficients between samples from different origins. In addition, the evidential value of such a match should be addressed by both showing high similarity (match) with the appointed origin (the grave for soil) in contrast to low similarity (mismatch) with alternative origins (backyard for soil).

13. PRESENTATION OF EVIDENCE

The overriding duty of those providing expert testimony is to the court and to the administration of justice. As such, evidence should be provided with honesty, integrity, objectivity and impartiality.
Evidence can be presented to the court either orally or in writing. Only information which is supported by the examinations carried out should be presented. Presentation of evidence should clearly state the results of any evaluation and interpretation of the examination.
Written reports should include all the relevant information in a clear, concise, structured and unambiguous manner as required by the relevant legal process. Written reports must be peer reviewed.
Expert witnesses should resist responding to questions that take them outside their field of expertise unless specifically directed by the court, and even then a declaration as to the limitations of their expertise should be made.

In general, the report contains the aim of the analysis, the methodology of analysis, the description of samples, the analytical results, the evaluation and the interpretation. If necessary, the limitations of sampling, methods and interpretation should be mentioned.

Independent from the methodology used, the reporting of the exact indication of the analysed genetic marker(s) could be a limitation in non-human DNA casework.
A standard reference sequence and an international nomenclature of STR-loci are not available for all species. In addition, the applied assays are not always published and reference to a publically available document may not be possible. As a result, the genetic markers that were analysed should be given in a general way (e.g. the sequenced gene, the number of loci), while more unambiguous details should be made available on request.
If a consultable publication is available, a reference sequence is internationally accepted (e.g. Kim sequence for dog mtDNA) or an international known reference sequence fragment (e.g. COI-gene fragment for CBOL) is analysed, this could be used as reference.

Each report has to conform to the general guidelines as mentioned above, however there are further issues which are particular to each of the three questions formulated in Chapter 5.

(i) What is it? Taxon identification

To answer this question in a scientifically correct manner it is important to state in the aim of the analysis the level of the taxonomic evaluation: i.e. (sub)species, (sub)genus or (sub)family. The

scientific name of the taxon must always be given in the report. However, if the taxon also has a commonly accepted popular name, this may be included alongside the scientific name to make the report more comprehensible for a non-expert.
The evaluation of the taxonomic classification can result in an inclusion if the specimen in question belongs to the taxon. If the specimen in question does not belong to the taxon, this will result in an exclusion.
If taxons cannot be distinguished by the locus used, this should be reported.

The consulted databases, their versions and their origin (internal and /or international) should be mentioned in the report. Also the time of consultation should be annotated as the content of databases is in a constant state of revision as data is added and removed over time. These modifications may influence the interpretation, and so variation in interpretation could also change with these revisions.
If a database was published and remains unchanged over time, references should be made to the appropriate scientific publication.

The interpretation of taxon identification could be influenced by the completeness of the consulted database. Paucities in the consulted database or phylogenetic tree should be mentioned. The effects of these deficiencies on the uncertainty of the interpretation should be stated.

If other scientific disciplines beyond DNA analysis were involved in the decision process of taxon identification, this should be clarified in the report. It could be useful to refer to a publication or external expert consulted.

(ii) To whom or to what does it belong? Individualization

STR-allele calling can be given as a number of repeats if an allelic ladder is available, or otherwise as a number of base pairs. If a comparison with other reports is necessary, the analysed reference sample or the allelic ladder used should be reported.

The individualisation can be considered an exclusion if there is no match between the questioned profile and the known profile. This is applicable for both haplotype sequences and nuclear DNA STR or SNP profiles.
In the event of match, and if an appropriate genotype database is available, the matching probability should be based on a LR calculation. For haplotyping, an alternative the number of hits of the particular sample sequence in a population database should be recorded.
The date and the version and size of the consulted database should be reported or referred to, since the interpretation could vary depending on the profiles present in the database.
The number of individuals in the database should be mentioned, in particular for haplotype interpretation as the frequency estimation for interpretation is directly based on the number of matches in the population database rather than the multiplication of frequencies of independent alleles, as for STR or SNP profiles from nuclear DNA.
If the results do not allow any decision to be made, the evaluation is inconclusive. A motivation on how this conclusion was drawn should be given.

Correction factors for kinship and/or inbreeding may also have an influence on the interpretation of results and therefore these should be reported, along with their motivation.

(iii) Where does it come from? Geographical or genealogical origin

Geographical/Population origin
An overview of the subpopulations and geographic areas considered could be stated in the report. If the locus/loci used are not sufficient to discriminate between some of the subpopulations, this has consequences for the interpretation and therefore these limitations in interpretation should be reported.

If statistical calculations were necessary for the interpretation, these should be reported to quantify the confidence of the assignment to a population/geographic origin.

The content of the consulted database should be described, including the geographic coverage and number of samples. Both considerations can have a marked influence on the interpretation of the results. Where gaps in the database limit the assignment of a sample to a certain subpopulation or geographical area, this should be mentioned in the report.

Biogeographical data of the species in question should be mentioned if these were used together with the DNA-analysis results in the interpretation. Reference should be made to any publications or experts consulted.

Genealogical origin
If mutation probabilities of STR alleles or the probability of a mutational event were used in relationship testing, this should be clearly reported.

If a match was found in the breeding lineage analysis an appropriate database should be used for interpretation. The size of this database has a direct influence on the match probability and should be provided. If the database has been published, a reference should be made. For a dynamic database the time of consultation should be mentioned if this is different from the time of reporting.

Inconclusive results should also be explained, for example describing the potential heteroplasmy which could be the source of variation between two generations.

14. HEALTH AND SAFETY

Working with biological traces can pose several health and safety risks for the investigator and can create cross-contamination issues inside laboratories. Some of these risks are specific to certain types of biological material (e.g. animal, plant, fungus), while others are related to laboratory practices in general.

14.1. Risks for humans and the environment

Animals and plants may carry diseases: this must be considered when handling their traces. If these are infectious diseases they may be passed on to other individuals of the same species or, in case of zoonoses, to humans. For zoonoses to spread from animals to humans, direct contact between an infected animal and a human is a necessity, or for a large number of infected animals to be present. Animal disease may also be of danger to humans when the causative agents of a human disease and an animal disease are present in a single reservoir and combine to form a new human pathogen. Also soil samples can carry pathogens. Animal diseases are considered a risk for food safety and human health, therefore many

domestic animals, both livestock and pets, are vaccinated against endemic diseases. This decreases the disease pressure in these animals, thereby decreasing the reservoir in which pathogens may reside. Livestock is generally not vaccinated against non-endemic diseases, but when such diseases are encountered, notification of authorities is mandatory and actions to prevent the spread of disease will be taken. Wild animals are typically infected by endemic diseases residing in soils and water or by contact with diseased livestock. Animal remains encountered in forensic investigations will generally not harbour diseases contagious for humans and good laboratory practices including, but not limited to, wearing appropriate laboratory coats, gloves and safety glasses should sufficiently protect the investigator. The risk of working with such remains cannot be considered higher than when working with human remains. However when an outbreak of any animal disease occurs, specific measures should be taken when investigating the remains of animals susceptible to such a disease (e.g. birds during a bird flu outbreak, goats, sheep or cattle during a Q-fever outbreak). Additional measures including the use of filtered facial masks, laminar flow cabinets etc. should be considered.

Animal remains (like human remains) can also harbour other pathogens, including fungi. The risks posed by such pathogens do not differ between human and animal remains and are minimized by the described good laboratory practices.

Practitioners should be aware that fragments from cultivated plants could be chemically treated. When analyses of DNA are required, laboratory procedures and/or chemicals will be used to degrade cell membranes (e.g. DNA-extraction), bind DNA (e.g. Ethidium bromide colouring) or degrade DNA (e.g. UV treatments, DNAse treatment). These procedures and chemicals will also affect the cells/DNA of the operators if the necessary precautions are not in place. The chemicals/procedures most likely to pose health risks for the investigator should be registered and actions to minimize the risks should be documented. All investigators should be familiar with any necessary precautions and follow good laboratory practices as described previously.

If laboratory waste is not disposed of in the correct manner, chemicals may also form an environmental threat and will be in breach of the law. If laboratories handling non-human DNA are part of larger biology or chemistry department, procedures for correct waste disposal may be copied from those departments. Alternatively, a laboratory wide waste management scheme may be in place. Individual laboratories which produce potentially hazardous waste must be aware of how to correctly dispose of this waste in accordance with the applicable local laws.

14.2. Risks for investigations

Contamination of samples by other samples from either the same or different cases is detrimental for a specific case, and also impacts negatively on the reputation of a laboratory and public perception of forensic investigations in general. To minimize the risk of (cross-) contamination potential causes should be identified and actions to minimize these risks should be enacted. For non-human DNA, the major concern is working with varying quantities of DNA: low quantities may be present in traces and pre-PCR samples while high quantities are present in reference samples and post-PCR. In addition to general laboratory cleanliness, separating procedures in time, operator and/or space should always be considered.

15. REFERENCES

QCC-BT-001, Guidance on the conduct of proficiency tests and collaborative exercises within ENFSI, version 001, 27/06/2014

QCC-VAL-002, Guidelines for the single laboratory Validation of Instrumental and Human Based Methods in Forensic Science, version 001, 10/11/2014

SWGWILD Standards and Guidelines v. 2.0 (2012).
http://www.wildlifeforensicscience.org/documents/2013/01/swgwild-standards_and_guidelines_2-0_12192012.pdf

Bär W, Brinkmann B, Budowle B, Carracedo A, Gill P, Holland M, Lincoln P, Mayr W, Morling N, Olaisen B, Schneider P, Tully G, Wilson M. DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing. Int. J. legal Med. 113 (2000): 193-196 and Forensic Sci. Int. 110 (2000): 79-85.

Budowle B, Garofano P, Hellmann A, Ketchum M, Kanthaswamy S, Parson W, Van Haeringen W, Fain S, Broad T. Recommendations for animal DNA forensic and identity testing. Int J Legal med 119 (2005): 295-302.

Dawnay N, Ogden R, Thorpe R, Pope L, Dawson D, McEwing R. A forensic STR profiling system for the Eurasian badger: a framework for developing profiling systems for wildlife species. FSI Genetics 2 (2008): 47-53.

Giampaoli S, Berti A, Di Maggio RM, Pilli E, Valentini A, Valeriani F, Gianfranceschi G, Barni F, Ripani L, Romano Spica V. The environmental biological signature; NGS profiling for forensic comparison of soils. FSI 240 (2014): 41-47.

Gill P, Sparkes R, Kimpton C. Development of guidelines to designate alleles using an STR multiplex system. For. Sci. Int. 89(3) (1997): 185-197

Hohof C, Schnöink K, Brinkmann B, Schneider PM. The GEDNAP proficiency testing system, June 2013. http://www.gednap.org/wp-content/uploads/2013/04/GEDNAP-Manual-E.pdf

Johnson RN, Wilson-Wilde L, Linacre A. Current and future directions of DNA in wildlife forensic science. FSI Genetics 10 (2014): 1-11

Koopmann W, Kuiper I, Klein-Geltink D, Sabatino G, Smulders M. Botanical DNA evidence in criminal cases: Knotgrass (Polygonum aviculare L.) as a model species. FSI Genetics 6 (2012): 366-374.

Linacre A, Gusmao L, Hecht W, Hellmann A, Mayr W, Parson W, Prinz M, Schneider P, Morling N. ISFG: Recommendations regarding the use of non-human (animal) DNA in forensic genetic investigations. FSI Genetics 5 (2011): 501-505.

Pereira F, Carneiro J, Matthiesen R, Van Asch B, Pinto N, Gusmao L, Amorim A. Identification of species by multiplex analysis of variable length sequences. Nucleic Acid Res. 38(22) (2010): e203

Tobe S, Linacre A. A multiplex assay to identify 18 European mammal species from mixtures using the mitochondrial cytochrome b gene. Electrophoresis 29 (2008): 340-347.

16. AMENDMENTS AGAINST PREVIOUS VERSION

Not applicable (first version)

Best Practice Manual for the Application of Molecular Methods for the Forensic Examination of Non-Human Biological Traces

ENFSI-BPM-APS-01

Version 01 - November 2015