European Commission

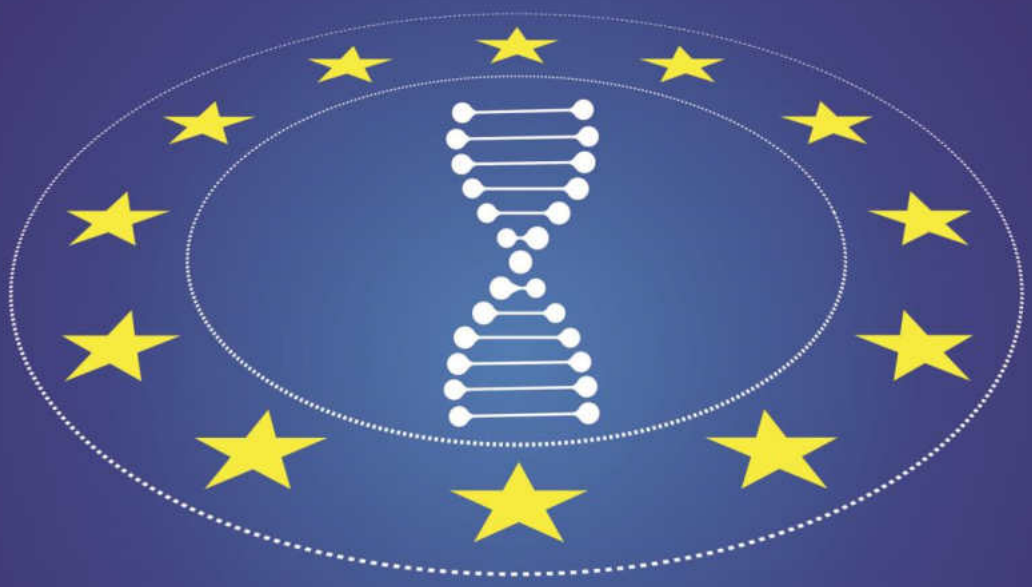
JRC SCIENCE FOR POLICY REPORT

Study on DNA Profiling Technology for its Implementation in the Central Schengen Information System

Administrative Arrangement JRC-34751

Angers, A.
Kagkli, D.M.
Oliva, L.
Petrillo, M.
Raffael, B.

2019



Joint Research Centre

EUR 29766 EN

Contact information
Name: Alexandre Angers
Address: Joint Research Centre, Via Enrico Fermi 2749, 21027 Ispra, Italy
Email: alexandre.angers@ec.europa.eu
Tel.: +39 0332 78 6539

JRC Science Hub
https://ec.europa.eu/jrc

JRC116742

EUR 29766 EN

PDF ISBN 978-92-76-07983-5 ISSN 1831-9424 doi:10.2760/13343

Luxembourg: Publications Office of the European Union, 2019

How to cite this report: Angers A, Kagkli DM, Oliva L, Petrillo M, Raffael B, *Study on DNA Profiling Technology for its Implementation in the Central Schengen Information System*, EUR 29766, Luxembourg: Publications Office of the European Union, 2019, ISBN 978-92-76-07983-5, doi:10.2760/13343, PUBSY No. JRC116742

Table of Contents

Abstract 4
Acknowledgements 5
Executive summary 6
Introduction 7
1 State-of-the art of the use of DNA markers in forensic science 9
1.1 Introduction to DNA 9
1.2 DNA as a forensic tool 9
1.3 DNA markers 10
1.3.1 Autosomal STRs 10
1.3.2 Sex chromosome STRs 12
1.3.2.1 X chromosome STRs 12
1.3.2.2 Y chromosome STRs 13
1.3.3 Amelogenin 13
1.3.4 Mitochondrial DNA 14
1.3.5 Other markers 15
1.3.5.1 SNPs 15
1.3.5.2 Microhaplotypes 15
1.3.5.3 InDels 15
1.3.5.4 LINEs/SINEs 16
2 Overview of the generation and use of DNA profiles to identify missing persons 17
2.1 General considerations 17
2.1.1 On the importance of the laboratory process 17
2.1.2 On the challenge of contamination 17
2.2 Step 1: Collection of biological samples 18
2.2.1 Selection of the DNA source 18
2.2.2 Sample collection 19
2.3 Step 2: From sample to profile 20
2.3.1 The DNA extraction step 20
2.3.2 The DNA quantification step 21
2.3.3 The DNA amplification step 21
2.3.4 On the use of massively parallel sequencing 22
2.4 Step 3: Evaluation of the genetic evidence 22
2.4.1 Direct comparison of two profiles 22
2.4.2 Kinship analysis 23
3 International Standards for DNA profiling methodologies 25
3.1 International Criminal Police Organization-INTERPOL 25

3.2 The Federal Bureau of Investigation 25
3.2.1 The Scientific Working Group of DNA Analysis Methods 26
3.3 The International Organization for Standardization 26
3.4 International expert groups 26
3.4.1 The European Network of Forensic Science Institutes 26
3.4.2 The International Society for Forensic Genetics 27
4 Current practices in established data repositories of DNA profiles 29
4.1 Differences between criminal and non-criminal DNA profile databases 29
4.2 Established missing persons databases 29
4.2.1 EU Member States 29
4.2.2 The International Commission on Missing Persons 30
4.2.3 The Spanish Phoenix program 31
4.2.4 The Committee on Missing persons in Cyprus 31
4.2.5 Missing Persons Bureau- UK 32
4.2.6 The United Nations human rights and the Red Cross suggestions 32
4.2.7 DNA-Prokids 32
4.3 Standard Sets of Short Tandem Repeats (STRs) 33
4.3.1 The CODIS Core Loci 33
4.3.2 The European Standard Set (ESS) 33
4.3.3 The INTERPOL Standard Set 34
4.4 Establishing family links in reference samples 34
5 Lessons learned from the exchange of DNA profiles under Prüm 37
5.1 The Prüm Decisions 37
5.2 Challenges in the implementation of the Prüm Decisions 37
5.2.1 Challenges related to the exchange of DNA profiles 38
5.2.2 Challenges related to the follow-up of requests made under Prüm 38
5.3 Future steps: the European Forensic Scientific Area 38
6 Evaluating the quality of a submitted DNA profile 40
6.1 Demonstrating the quality of the process that generated the profiles 40
6.2 Verifying the quality of the experiment that generated the profile 41
6.3 Checking for "hints" of low quality in the final text file 42
6.3.1 Contamination 42
6.3.2 Low quantity or poor quality of the purified DNA used to generate the profile: dropped alleles. 43
6.3.3 Consistency of the marker values with population information 43
6.3.4 Mitochondrial DNA sequence: the FASTQ format 45
6.4 Standardised formats for DNA profiles 46
6.4.1 Prüm 46
6.4.2 CODIS 47

6.4.3 ISO/IEC 19794-14 49
6.4.4 Verifying internal consistency 49
7 Conclusions and Recommendations 52
7.1 DNA profiles - data 52
7.2 DNA profiles - metadata 53
7.3 DNA profiles - format 53
7.4 Quality of the submitted DNA profiles 54
7.5 General comments 55
7.6 Future steps 55
8 References 56
List of abbreviations and definitions 62
List of figures 64
List of tables 65
Annex 66

Abstract

In 2018, Regulation (EU) 2018/1862 added the possibility to introduce DNA profiles in alerts related to missing persons, in the Schengen Information System (SIS), in order to contribute to their identification. The present report describes the state-of-the-art for the generation and use of DNA profiles for individualisation purposes. The objective is to inform on the type of data that compose a DNA profile, and to propose ways to evaluate its quality.

The report starts with an overview of the different types of DNA markers that are currently used in DNA-based forensics procedures, and a description of the processes involved in the generation and use of DNA profiles. It then provides an overview of important working groups and organisations that are active in establishing standards and best practices in the field, and examples of existing databases developed for the identification of missing persons. The practices and experiences of exchanging DNA profiles between Member States in the context of the Prüm Regulation are also discussed, highlighting the instances where the lessons learnt could be relevant for the SIS. The report concludes with an overview of the different levels of quality checks that can be performed on DNA profiles prior to their insertion in a database.

Acknowledgements

This report was carried out by members of the DG JRC "Knowledge for Health and Consumer Safety" unit of Directorate F – Health, Consumers and Reference Materials.

The authors would like to express their gratitude to the following persons that took the time to receive us in their premises and discuss with us during the fulfilment of the study:

- Dr Susan Hitchin, Coordinator of the DNA Unit of INTERPOL
- Dr Thomas Parsons and his group (in particular René Huel, Adnan Rizvic and Andreas Kleiser) from the International Commission on Missing Persons.

We would like to thank, for accepting to review the first draft of this report and for greatly improving it with their comments and corrections:

- Ms Līga Peisniece, Deputy Director, Forensic Service Department, State Police of Ministry of Interior of Republic of Latvia.
- Dr Jiří Drábek, Professor, Palacky University Olomouc, Czech Republic.

We are also grateful to our JRC colleagues from the "Cyber and Digital Citizens' Security" Unit E.3, namely Laurent Beslay, Pasquale Ferrara, Javier Galbally and Rudolf Haraksim for taking the time to read and comment.

Executive summary

This report is a DG JRC study regarding the state-of-the-art of DNA profiling technologies for the introduction of DNA profiles in alerts within the Central Schengen Information System (CS-SIS). The study was carried out for DG HOME via an Administrative Arrangement.

Policy context

Created as a compensatory measure for the abolition of internal border checks within the Schengen area, the SIS was established with two intentions: to contribute to police and law enforcement cooperation between the Member States and to support external border control. In its first generation the SIS was the first large-scale IT system launched by the EU Member States in 1995. It was followed by EURODAC (asylum seekers' database) in 2003 and the Visa Information System (VIS) in 2011.

The second-generation of the SIS entered into operation on 9 April 2013. The CS-SIS offers the possibility to store biometric data in alerts related to persons. In addition to alphanumeric data, alerts related to persons should contain fingerprints as well as facial image of the subject of the alert, whenever these are available. On 28th November 2018, Regulation (EU) 2018/1862 added (among other changes) the possibility to also introduce DNA profiles for a specific subset of alerts, those related to missing persons who need to be placed under protection either for their own protection in order to prevent a threat to public order or public security (point (a) of Article 32(1)). Even then, DNA profiles may only be added when photographs, facial images or dactyloscopic data are not available or not suitable for identification.

In support of this newly adopted 2018 Regulation, the objective of the present DG JRC study is to describe the state-of-the-art of how DNA profiles are generated and used for the purpose of identification, with a view to explain the type of data (content and format) that compose DNA profiles and should thus be stored by the CS-SIS when included in an alert. The report also describes different levels of quality checks that are performed on DNA profiles prior to their insertion in a database.

Key conclusions

DNA profiling relies on the characterisation of a minute fraction of a given persons' DNA at very specific regions of the genome. The most commonly used markers in this context are genomic regions called Short Tandem Repeats (STRs), and can be expressed as the number of repeats (as a number) present for each marker for a given individual. Standard sets of these markers have been designed, according to specific criteria, and consist of about 10-20 specific markers. Since a profile can be expressed as alphanumeric text, different XML-based schema have been proposed to express (and exchange) DNA profile information.

The generation of a DNA profile from a biological sample is a complicated and time-consuming procedure that needs sophisticated equipment and fully trained lab specialists. Efforts have been made to produce standards and best practices in this process in order to minimise problems related to, for example, contamination of the sample and poor quality of the purified DNA. In some cases, these issues can be inferred from the information present in the DNA profile itself.

Introduction

Biometric identifiers are the distinctive, measurable, biological and behavioural characteristics that can be used to describe individuals. Once transformed or captured in a computer-readable format, biometric data can allow high-confidence identification of individuals and are thus crucial in forensic sciences.

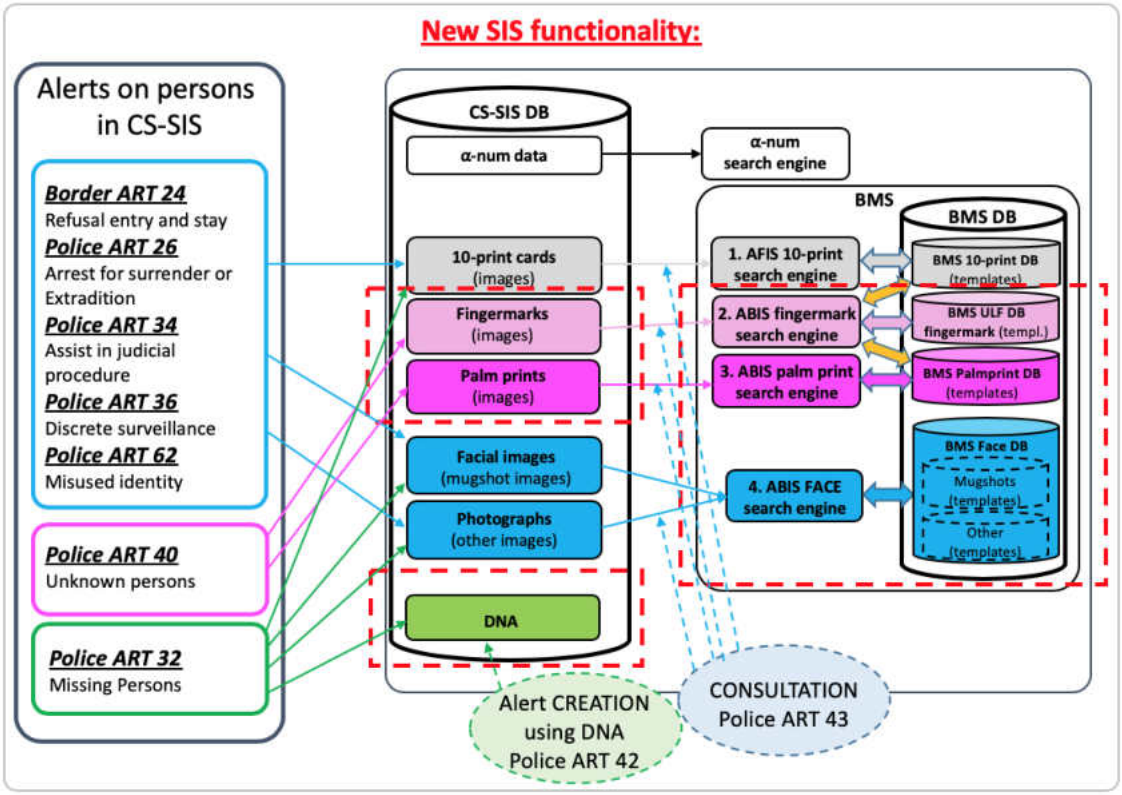
The possibility to share biometric data from wanted persons, persons who may not have the right to enter or stay in the EU, as well as missing persons, is an important component of a cooperation tool such as the Schengen Information System (SIS). The possibility to search the alerts stored in the Central System of the SIS (CS-SIS) with fingerprints for the purpose of identification of a person had already been set out in the decisions pertaining to the establishment, operation and use of the second generation SIS [1,2].

On 28th November 2018, Regulation (EU) 2018/1862 added (among other changes) the possibility to introduce DNA profiles in an alert related to a missing person (Article 32(1)) in order to contribute to their identification [3]. **Figure 1** below gives an overview of all the new functionalities, including the use of DNA profiles.

In particular, Article 42 states:

"(3) A DNA profile may only be added to alerts in the situations provided for in point (a) of Article 32(1), only following a quality check to ascertain whether the minimum data quality standards and technical specifications have been met and only where photographs, facial images or dactyloscopic data are not available or not suitable for identification. The DNA profiles of persons who are direct ascendants, descendants or siblings of the subject of the alert may be added to the alert provided that those persons give their explicit consent. Where a DNA profile is added to an alert, that profile shall contain the minimum information strictly necessary for the identification of the missing person."

Figure 1. Present functionality of CS-SIS in the scope of the recently adopted regulation [3].



Source: JRC, 2018

This JRC report presents an assessment of the state of the art on the use of DNA profiles for individualisation purposes. This work was performed in the context of an Administrative Agreement between DG JRC and DG HOME regarding the integration or deployment of biometric capabilities within the SIS. This report is part of a set of three Science for Policy Reports, with the other two covering, respectively, fingerprint and face identification technologies.

The objective is to inform on the type of data expected to be submitted to the CS-SIS, and to propose ways to evaluate their quality in line with the text of Article 42 quoted above.

The report starts with an overview (Section 1) of the different types of DNA markers that are currently used in DNA-based forensics procedures. Section 2 provides an overview of the field and laboratory work involved in the generation of DNA profiles, emphasising the factors that can influence the quality of the final profiles. The subsequent Sections 3 and 4 provide an overview of important working groups and organisations that are active in establishing standards and best practices in the field, and of existing databases developed for the identification of missing persons. Then, the practices and experiences of exchanging DNA profiles between Member States in the context of the Prüm Regulation are summarised in Section 5, highlighting the instances where the lessons learned could be relevant for the SIS. Finally, Section 6 proposes different levels of quality checks to be performed on DNA profiles prior to their insertion in the databases.

The conclusions and recommendations from the different Sections of the report are summarised in Section 7, and focus on two main elements:

- What is a DNA profile (data, metadata and format)?
- How can its "quality" be evaluated?

1 State-of-the art of the use of DNA markers in forensic science

This Section provides a brief introduction on the biology of DNA and the types of DNA markers that are being used when generating profiles for forensics purposes.

1.1 Introduction to DNA

Deoxyribonucleic acid (DNA) is a molecule composed of a string of small building blocks called "nucleotides". For most organisms on Earth (including humans), there are only four canonical blocks, labelled "A", "C", "G" and "T" according to the initials of their chemical names. DNA is organised in the well-known "double helix" where two strands twist around each other, usually in the right-handed manner, to form a "ladder"; however, when writing a DNA sequence, only one strand is described; the second strand is omitted for simplicity because A in one strand is matched (bound to) with T in the other strand and C in one strand is matched with G in the other strand and vice versa. Because of a chemical asymmetry in the bonds between the bases (one strand runs from 5´ carbon atom of deoxyribose to 3´ one while the other runs from 3´ to 5´), a DNA sequence is directional, e.g. the sequence "AAGGTTCC" is distinct from the sequence "CCTTGGAA".

Within a human cell, DNA is found in two main locations: the nucleus and the mitochondria. Outside of the human cell, some DNA may be found in circulating form in the blood (usually in higher amounts in pregnancy and cancer).

In the nucleus, the DNA is organised in 23 distinct chains called chromosomes of various length (for example, chromosome 1 is a chain of about 250 million nucleotides linked together, while chromosome 22 is about 50 million nucleotides long). Barring rare exceptions (e.g. gametes), every chromosome is present in two copies: one received from the mother and one from the father. One of the pairs represents the sex chromosomes "allosomes" (XX for females, XY for males), while the other 22 chromosomes are called "autosomes" and are numbered 1 to 22.

In the mitochondria, the DNA is circular, i.e. the two ends are linked to each other. It is relatively short, about 16 000 bases. Importantly, the mitochondria (and their DNA) are inherited as they are from the mother, without a contribution from the father. This has forensic consequences, described below.

1.2 DNA as a forensic tool

There are a few biological and technological reasons that make DNA an important source of biometric data. These include:

- **DNA is present and identical in all cells of an individual.** DNA can be extracted from different biological samples (blood, saliva, hair follicles, etc.) and the DNA markers generated from any of these samples for the same individual will show the same values. There are some exceptions to this; for example, red blood cells don't have a nucleus, so no nuclear DNA, and gametes only have one copy of each nuclear chromosome.
- **DNA is inherited.** Biologically, DNA is the carrier of genetic instructions for all the biological functions of the cells, and the medium by which these instructions are transmitted to the next generations. DNA markers can thus, in addition to providing evidence of unprecedented strength for or against the individualization of the profile, support or suppress the scenario of familial links between two persons (unlike, for example, fingerprints).
- **DNA is chemically stable.** The bonds linking the individual bases to each other are stable (more than bonds between amino acids in proteins), and DNA molecules are known to be stable over time and to survive exposure to heat and cold. As extreme examples, DNA is used for the regulatory control of highly processed food and food

supplements [4], and DNA was recently used to determine the biological sex of a 4000-year-old Egyptian mummy [5].

- **DNA can be efficiently amplified and/or sequenced.** The first techniques using DNA for forensics required relatively large amounts of DNA material - in the micrograms range. Since then, a technique called "Polymerase Chain Reaction" (PCR) has been developed and perfected, allowing the exponential amplification of specific DNA regions in a sample [6], such as the regions used as forensic markers. With this technology, current protocols suggest the use of less than 1 nanogram of DNA in the reaction, which is easy to obtain from most biological sources (see Section 2). In addition, technologies that allow the sequencing of DNA (i.e. reading the sequence of bases, for example in the regions used as forensic markers) have also greatly improved in accuracy, throughput and price.

On the other hand, DNA profiling is a complicated and time-consuming procedure that needs sophisticated equipment and fully trained laboratory specialists. DNA can be vulnerable to mistakes or mishandling, and DNA profiling, like any other technique in forensic science, is susceptible to human error. If an error occurs during the extensive process of sampling or processing (see Section 2), then the results may not be accurate. In particular, contamination during sample collection or analysis can dramatically decrease the accuracy and reliability of the resulting DNA profile.

1.3 DNA markers

Because of its important role in all the biological functions of the different cells in the body, the overwhelming majority of the DNA sequences of two individuals will be identical, thus not useful as biometric data. The exact amount of variation among individuals is difficult to establish, and vary according to the way of calculating it, but it has been reported to be only in the range of 1-2% for nuclear DNA [7]. Even then, this limited amount of variation is much more information than needed to provide the needed evidence with sufficient strength (the whole nuclear genome is more than 3 billion bases), therefore a restricted set of specific markers have been selected as sufficiently informative for this purpose, described in this Section.

1.3.1 Autosomal STRs

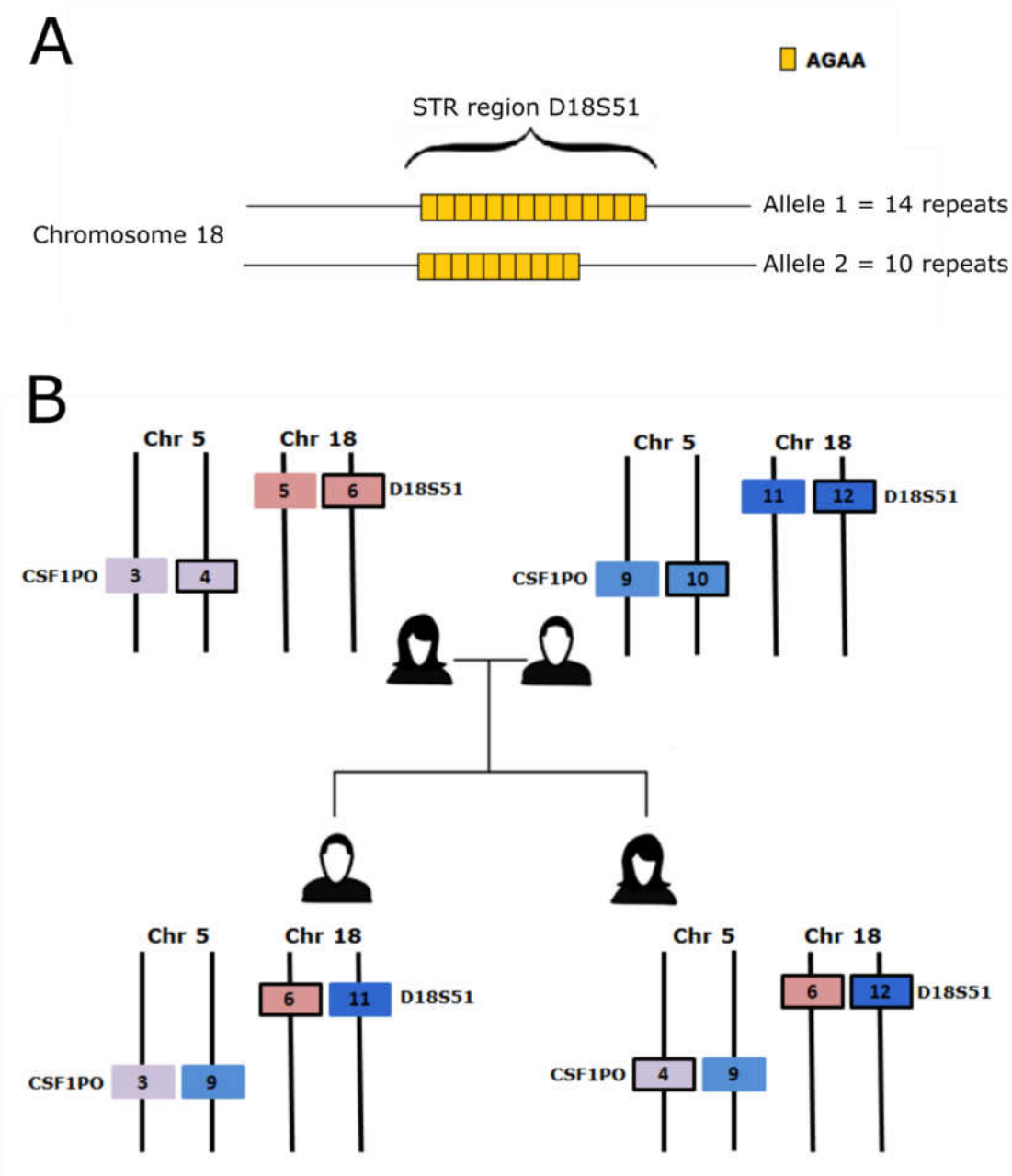
Short Tandem Repeats (STRs) are tracts of tandemly repeated DNA motifs that range in length from three to five bases, found at different places in the genome. They are called "autosomal STRs" when they are found in one of the 22 autosomes in the nucleus, and are the most commonly used DNA markers for DNA profiling.

STRs can occur in regions that do not have any specific biological function, meaning that evolution does not actively select for or against the mutations that accumulate over the generations, when DNA replication makes the mistake of adding or removing one of the repeats. This allows variability in these locations (loci) in the form of the number of repeats present in the genomes of different individuals that can be used for DNA profiling.

As an example, the STR locus called "D18S51" is found on the chromosome 18 and consists of many repeats of the motif "AGAA" [8]. Population studies identified occurrences in the number of repeats ranging from 7 to 40 in different individuals [9]. Because chromosome 18, like all autosomes, is found in two copies in each cell (one from the father, one from the mother), each individual will carry two instances (alleles) of this locus. Figure 2A illustrates the example of a person having 10 repeats on one of the two alleles, and 14 on the other.

A study analysing 1 000 unrelated individuals from the Lodz region of Poland showed that the frequencies of these two alleles in this set were 0.0095 and 0.1605, respectively [10], meaning that within that context, two samples showing these same values (10 and 14) for this marker would provide some support to the scenario of the same person origin of the two samples, whatever the prior odds of this scenario are.

Figure 2. Visual illustration of the biology and inheritance of STRs. (A) Marker D18S51 consists of repeats of the sequence AGAA (yellow rectangle). In this example, one allele contains 10 of these repeats, the second allele 14. (B) For each chromosome, parents contribute one of their two copies to their offspring, selected randomly in the genesis of the gametes. This schema illustrates one of the possible combinations (for two markers), given by the parents to their children.



Source: JRC, 2019

Of course, this is not sufficient to conclude anything in a forensic context, based on one item of evidence, even if the evidence is as strong as DNA. Authorized agencies (e.g. court) achieve identification by taking into account scenarios compared, prior odds, strength of evidence (expressed as a likelihood ratio that take into account a larger number of markers), and threshold for identification.

Likelihood ratio for the scenario of the same source of two samples versus scenario of two sources will depend on 1) the specifics of the second scenario (e.g. whether the

source of scene sample is unknown, unrelated to a man or some brother of the alleged man), 2) the number of loci used, 3) the number of possible values observed in the population for each of the loci and 4) the proportions in a population of each of the values for each of the loci (discussed further in Section 2.4). It is also crucial that the same markers are used in the two profiles to be compared; otherwise drawing conclusions would be (almost) impossible[1]. For this reason, sets of standard loci have been agreed through international harmonisation efforts, described in Section 4.3. This also requires coordinated effort to obtain and disseminate the empirical information for the possible markers values and their proportions in different populations, together with their mutation frequencies.

Because the two alleles of each marker are inherited from the parents, autosomal STRs can also be used to identify familial links between individuals. For each STR marker, one of the values for an individual should match one of the two values of the same marker of his/her biological father, the other should match one of the two values of the same marker of his/her biological mother. How these get distributed, at conception, is unpredictable. An example is shown in Figure 2B for two markers (one on chromosome 5, the other on chromosome 18) and two siblings. The calculations of likelihood ratio supporting kinship, given a set of patterns (profiles in person within a supposed pedigree) is thus more complex than matching two samples to the same individual, and may require the use of a greater number of markers, see Section 2.4.2 [12].

1.3.2 Sex chromosome STRs

Sets of STR markers have also been identified on the two sex chromosomes (X and Y). In many "standard" cases, the use of these markers would be superfluous, as autosomal STRs would be sufficient to provide sufficiently high evidence of the same source of two samples. However, when depending on kin information, there are instances where these markers are necessary to complement autosomal STRs. These considerations are particularly important in the context of the SIS, as Regulation (EU) 2018/1862 states that "*Where a DNA profile is added to an alert, that profile shall contain the minimum information strictly necessary for the identification of the missing person.*" (Article 42(3)).

1.3.2.1 X chromosome STRs

The markers on the X chromosome are similar to the autosomal markers, with the distinction that males only have one X chromosome, which is always inherited from their mother (the second of the pair is the Y chromosome inherited from their father, see below).

X-STR markers are useful in the identification of victim remains in case of mass disasters, in particular to ascertain specific kin relationships that are not possible to distinguish using only autosomal STRs. The most common of these cases is the fact that autosomal markers cannot distinguish (i.e. would give the exact same likelihoods for) the following relationships between the donors of two DNA profiles: grandparent–grandchild, uncle–niece and half-siblings, as the likelihood ratio comparing any pair of hypotheses will be unity [13]. In these cases, there are specific combinations that can be solved using X-STR markers, which are described in detail by Pinto et al. [14].

For the identification of missing persons *per se* who are subject to an Article 32(1) alert, the use of X-STRs, although not common, has been shown to be useful in cases where the only available reference DNA profiles were the missing person's siblings (or half-siblings), complementing the autosomal STRs to reach the acceptable likelihood ratio for identification [15].

[1] As sometimes markers are "linked" to each other, disjoint sets of forensic markers may provide some amount of evidence in some circumstances, see [11]

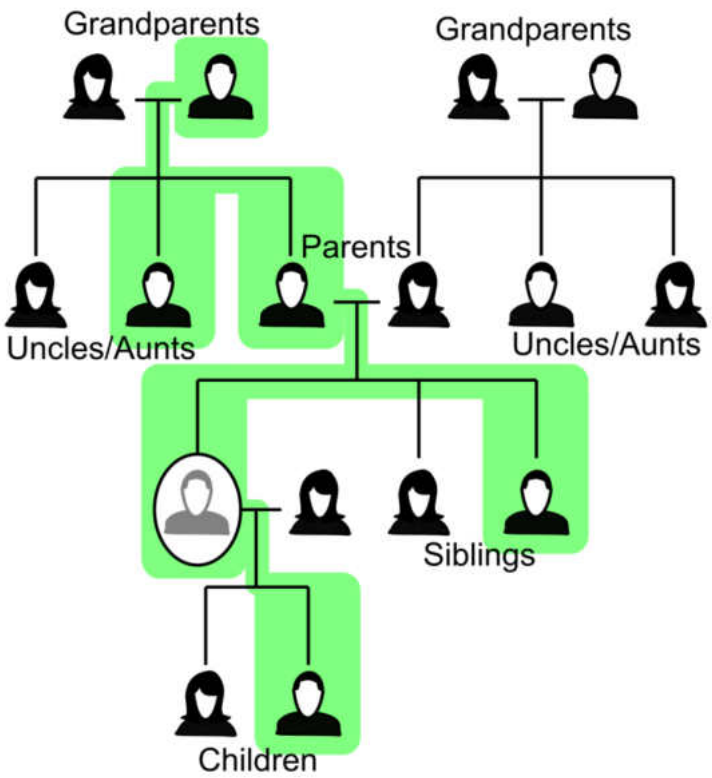
1.3.2.2 Y chromosome STRs

The Y chromosomes are specific to males. They are transmitted directly from father to son.

In forensics, Y-STRs are most useful for crime investigations, allowing identification of paternal lineage or even inference of paternal bio-geographic ancestry of unknown male donors. They are particularly useful in the case of a mixed sample, when Y-STRs can generate a male perpetrator DNA profile without risks of contamination with the profile of the victim, if female [16].

Commercial Y-STR kits are not suitable, on their own, for the identification of a male individual, because male relatives typically share identical patterns for these markers. However, since Y-STRs can be used to identify groups of male relatives belonging to the same paternal lineage, they are suitable (and sometimes necessary) to identify a missing person, if they are male and only the profiles of distant male relatives are available (for example, an uncle on the male line, see Figure 3). Y-STRs are not informative in the cases where the missing person is female.

Figure 3. Relatives from which Y-STR information could be informative to identify a missing person (circled), if - and only if - the missing person is male.



Source: JRC, 2019

1.3.3 Amelogenin

Amelogenin is a gene present on both the X and Y chromosomes [17]. Though it is not an STR, there is a length polymorphism, i.e., the version on the X chromosome has a 6-base deletion compared to the Y chromosome version. The two versions (X and Y) can thus be differentiated in the laboratory using the same technique as for STRs (capillary electrophoresis, distinguishing length of amplicon), allowing the use of this marker to

identify the sex of the sample donor [18]. Since, for missing persons alerts, both the sex of the missing person and of the other potential DNA profile donors are known, this marker is not particularly useful; however, it is often part of the standard DNA profiling kits and may be, because of that, present in the submitted profile.

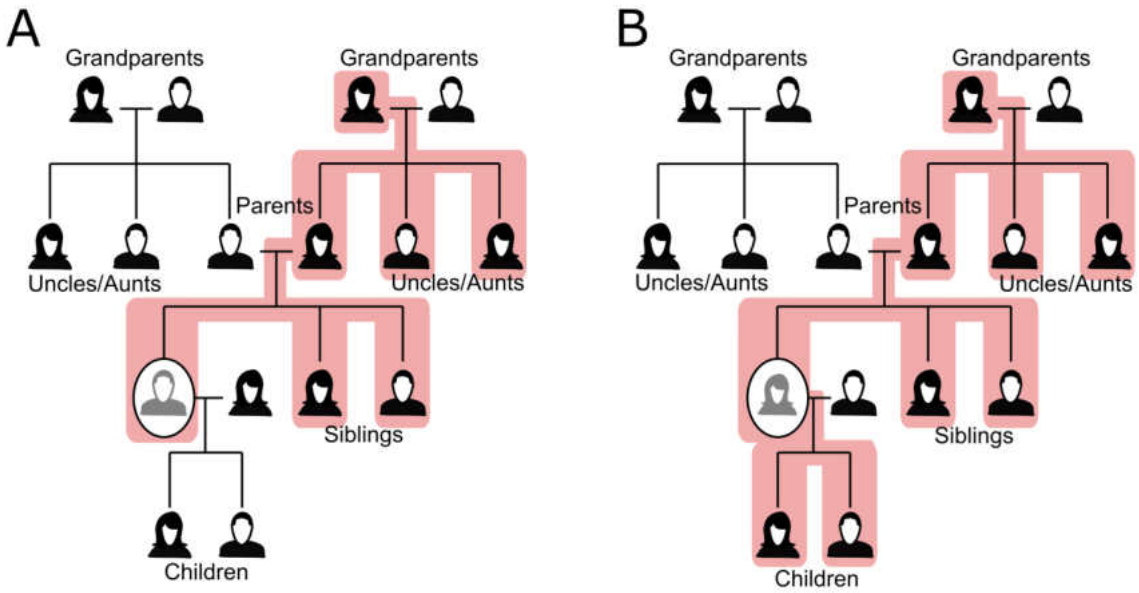
1.3.4 Mitochondrial DNA

Mitochondrial DNA (mtDNA) is a circular genome of about 16.5 kb that is located inside the mitochondria [19]. About 1000 mitochondria are present in most cells, meaning that mtDNA markers are present in two or three orders of magnitude more than their nuclear counterparts. Thus, mtDNA can more readily be recovered from highly degraded forensic and archaeological specimens. Like sex chromosomes markers, analysis of mtDNA is a powerful tool for identifying family links as a complement to the analysis of autosomal STRs.

In forensic science, in particular for missing persons and unidentified remains, two specific regions of the mitochondrial DNA sequence, the hypervariable regions (HVI and HVII) [20], are used. Because mitochondria are transmitted only from mothers to their children (with no contribution from the father)[2], mtDNA can be used to support or suppress scenarios of relationships on the maternal lineage between individuals [22], as illustrated in Figure 4.

It should be noted that mtDNA is usually used in countries or events with large databases and relatively large number of unidentified bodies. Also, mitochondria do not have a repair mechanism, so mutations are spread clonally and can generate mixtures of two (or more) mitochondrial genotypes in the tissue, a phenomenon called heteroplasmy.

Figure 4. Relatives from which mtDNA information could be informative to identify a missing person (circled), if the missing person is male (A) or female (B).



Source: JRC, 2019

(2) It should be noted that, although this is the general rule, some exceptions have recently been reported, see [21]

1.3.5 Other markers

The previous Sections described the DNA markers that are generally used for routine forensic purposes and are the main components of current DNA profile databases, both criminal and non-criminal (see Section 4).

This field remains very active in terms of research and innovation, and the use of new markers that have advantages in specific applications has been described. They are shortly summarised below, even though it is not expected that they will be included in profiles submitted to the CS-SIS in the near future.

1.3.5.1 SNPs

A single-nucleotide polymorphism (SNP) is a common genetic variation that occurs when a single nucleotide, in a specific position in a genome, is different among members of the same species (assuming a frequency >1% in the population); in every individual there are about 5 million SNPs.

The SNPs can be considered a tool in forensic genetic when the recoverable DNA is too degraded (in the case of human remains or mass disasters) to use STRs because the analysis can be performed on shorter fragments of DNA [23]. However, individual SNPs have much less variation than STRs (at best, 4 different possibilities in the population, "A", "G", "T" or "C") many more SNPs markers are needed (about 50-100) to have the same information level that commercial STR multiplex kits with ≥15 loci provide [24].

SNPs can be generally divided in at least 4 groups in the forensics field:

- Identity-testing SNPs (that help to individualize);
- Lineage informative SNPs (shared between close kin over several generations; mitochondrial, Y chromosome and mini/microhaplotypes –see below- may be considered here);
- Ancestry informative SNPs (extension of lineage informative SNPs over more generations; they give an indication, for example, of the biogeographic or ethnic origin of the person);
- Phenotype informative SNPs (that predicts other biometric characteristics of the person, e.g. hair and eye colour - or even the shape of the face, see http://www.visage-h2020.eu/ or [25]).

It is clear that some of the polymorphism groups can cause potential issues with respect to individual privacy [24] and go beyond the scope of confirming the identity of missing persons.

1.3.5.2 Microhaplotypes

Microhaplotypes are small regions of genome (< 300 nucleotides) that contain two or more SNPs very close to each other, presenting multiple allelic combinations ("haplotypes") [26]. Currently, around 130 have been identified [22], and these are expected to be used for the individualization, biogeographic ancestry inference, mixture deconvolution, relationship testing, and identification of missing persons [26].

1.3.5.3 InDels

Insertion-deletion polymorphisms (InDels) are genome variations due to insertion or deletion of short DNA segments (1-50 bp) on the chromosomes [27]. InDels are one of the most common polymorphisms in humans and can be a good alternative to mini-STRs that can be analysed by capillary electrophoresis with the same simple end-labelled PCR primer methods as STRs [22]. These DNA markers have been taken into consideration in the forensic community, since, like SNPs, InDels can be detected in degraded DNA samples, present a low mutation rate, and permit a high multiplexing capability [27]. Fondevila et al. [28] suggest that the typing of about 60 InDels can genetically identify human remains.

1.3.5.4 LINEs/SINEs

The Long Interspersed Elements (LINEs) and the Short Interspersed Elements (SINEs) are retrotransposons, fragments of DNA that can transcribe themselves into an intermediate RNA and then replicate themselves in sparse regions of the genome [29].

Both LINEs and SINEs have found applications as DNA markers for forensics [22]. The LINEs are long sequences of DNA able to encode a gene that allows the copying and transposing of both themselves and SINEs. The SINEs (80–400 bp in length) cover about 40% of the whole genome; the most common SINEs in humans belong to the family of the Alu sequences (short DNA sequences which are recognized by the *Arthrobacter luteus* (Alu) restriction endonuclease) which, because they are present in a very high number of copies (one million per human genome) and are identical by descent only without parallel independent insertions, make them excellent for the human identification and kinship testing of very degraded samples [22].

Section 1. Summary of key concepts

- The DNA profiles attached to missing persons alerts in the CS-SIS should be composed of STRs, described as the name of the markers attached to their values (number of repeats on each of the two alleles).
- The use of other markers, such as SNPs, is not expected at this stage, due to the state of the art of the harmonised implementation of this technology.
- For the missing person, the profile should be composed exclusively of autosomal STRs.
- For the profiles of family members, the use of other markers (Y or X chromosomes STRs, mtDNA) should be allowed, if they make sense based on their inheritance pattern relative to the missing person.

2 Overview of the generation and use of DNA profiles to identify missing persons

The process of generating DNA profiles to assist in the identification of missing persons consists of two main steps:

Step 1: Collection of biological samples from which the profiles will be generated. These can include intimate items of the missing person and family relatives ("reference samples").

Step 2: DNA analysis from these samples, which includes a series of laboratory phases like DNA extraction, quantification, amplification, and DNA profile determination in a machine-readable format (i.e. as biometric data).

Once a candidate is found for the missing person, either alive or as human remains, a DNA profile has to be generated from that sample ("the test sample"), using the same two steps.

To ensure the quality of the generated profile, it is very important to define strict rules and procedures to harmonise the steps for the production of genetic profiles from the starting biological material. The procedures to process the test samples should be adopted to reflect and accommodate the requirements of those used to process the reference samples. These rules have been reported and described in a set of articles and guidance documents [30–37], and their main conclusions are summarised in this Section.

Then, a third step is performed:

Step 3: Comparison of genetic profiles between the examined sample and the reference sample(s) and calculation of the likelihood ratio.

2.1 General considerations

2.1.1 On the importance of the laboratory process

When dealing with DNA samples, it is very important to take into account and strictly follow agreed protocols. In particular, many factors can easily lead to contamination of samples, such as inappropriate handling of the sample, disposables that do not fit ISO 18385:2016 [38] or damaged storage containers. The more steps and individuals involved in the DNA collection, extraction and amplification process, the higher the probability of contamination.

For this reason, it is a good practice to split the operational steps of a forensic DNA laboratory into four groups: pre-laboratory, pre-PCR, PCR, and post laboratory. The pre-laboratory steps regard case assessment and collection of the sample, while the pre-PCR steps include inspections, DNA extraction, and DNA quantification, (by fluorimetry, if used) PCR steps include DNA amplification, electrophoresis and typing. Finally, the post laboratory steps are mainly the interpretation of results, updating databases, and reporting. It is thus highly recommended to execute these steps in different areas to avoid contamination and have in place rigid and strict protocols to ensure a smooth transition from one step to another.

The competence of the staff, the proper calibration and maintenance of the equipment, the proper validation and quality management of procedures and protocols are crucial for the quality of the generated DNA profile. Request for laboratory compliance with standard ISO practices is highly recommended.

2.1.2 On the challenge of contamination

Contamination is a challenge that must be carefully monitored during the whole process in which DNA is managed, in order to correctly obtain a good quality representative DNA extract to be analysed. It may occur in laboratories, or during trace gathering at the scene.

Below, a set of broad and pragmatic recommendations is reported:

- During the collection phase, samples to be analysed should ideally be recovered as soon as possible, with appropriate equipment (e.g. shoe covers, lab coats with hoods, gloves and facemasks), to prevent introducing operators' DNA.
- During laboratory analysis, it is fundamental not to introduce the operator's DNA, often already present in already used reagents or not well cleaned lab surfaces. For this reason, DNA extractions, including extraction blanks, should be carried out in dedicated extraction hoods, and multiple no-template controls (i.e. a reaction in which there is no DNA) should be included in all DNA amplification steps to monitor contamination.
- In the laboratory where DNA extraction is performed, DNA-free, air-tight containers should be used to avoid contamination of DNA from the surrounding environment.
- It is fundamental to maintain the rigorous use of controls to monitor contamination at all steps of the experimental process.
- The no-template controls should be processed and analysed in parallel to all samples. In case positive results are produced from these controls, the source of contamination should be identified and removed. It should be considered that negative no-template controls do not exclude the possibility of low level contamination within individual reactions.
- Positive controls should be carefully used, as they are a potential source of cross-contamination within the experiment itself. For example, it is better to use positive controls to check the reagents, but at low copies and possibly prepared after the preparation of the samples to be analysed.

The above list is not exhaustive and for a whole and comprehensive picture, the ENSFI "DNA Contamination prevention guidelines"[3] provides recommendations for the minimum requirements for the laboratory layout and analysis (test) design to prevent the occurrence of contamination.

For those cases of missing persons in which the material to be analysed is a biological remain of a crime scene, additional provisions are included in the Anti-Contamination Strategy [39]. This strategy is designed around the known circumstances of the investigation, properly documented and effectively communicated to all relevant involved staff. The factors to be considered and reported in this strategy include:

- Prior to scene attendance, which regards all individuals including investigators, witnesses, suspects or other members of the public that might have contaminated the scene beforehand;
- Identification of risky environmental factors, e.g. hot conditions, linking routes;
- Staff deployment, e.g. avoiding involvement of the same personnel, vehicles or equipment that have been used in a linked scene; scene assessment, which includes sufficient and positioned appropriately cordons and scene protection;
- Identification of contamination risks between different parts of the same scene; avoiding the use of dogs before collection of trace DNA, which may introduce DNA from outside the scene.

2.2 Step 1: Collection of biological samples

2.2.1 Selection of the DNA source

As described in Section 1, genetic analysis involves the characterisation and identification of sequence targets on all major classes of genetic material, including those on

([3]) http://enfsi.eu/wp-content/uploads/2017/09/DNA-contamination-prevention-guidelines-v2.pdf

autosomal (non-sex) chromosomes, X and Y (allosomal, sex) chromosomes of the nuclear genome, as well as the mitochondrial genome. For all of them, in current forensic DNA analysis procedures, it is necessary to extract and purify DNA from the biological sample, in order to obtain the DNA molecule in sufficient quantity and purity that will allow the markers it contains to be characterized.

Well established and routine methods for extraction and purification of DNA from biological samples are available. They allow extracting DNA from different *in vivo* sources (blood, hair, saliva, semen, etc.) as well as from post-mortem remains, although for the latter it may be more problematic. In fact, despite its relative stability in proper conditions, DNA may decay very rapidly depending on the environmental conditions. For example, in hot and humid climates, DNA may be become rapidly fragmented and chemically modified, or metabolised by bacteria and moulds.

The quantity of DNA recovered varies according to both the sample type and its condition. Table 1 summarises the amount of DNA that is usually recovered from different sources.

Table 1. Maximum amounts of typically recoverable DNA from the various biological materials used in forensic sciences

Source	**Max Recoverable DNA**
Blood	20,000–40,000 ng/mL
Semen	150,000–300,000 ng/mL
Plucked hair roots	750 ng/plucked hair roots
Shed hairs	1–12 ng/hair
Bones	3–10 ng/mg bone
Urine	1-20 ng/mL
Saliva	1,00-10,000 ng/mL

Source: Summarised from [35]

Table 1 shows that the best and most convenient sources are from blood and saliva. Samples taken directly from living persons (i.e. relatives of the missing person) would be the preferred sources for the generation of their profiles, if available. For the missing person themselves, leftover samples (e.g. remains such as blood stains and intimate objects with DNA traces such as toothbrushes, combs and cigarettes) could also be used providing that there is sufficient confidence regarding their pristine origin.

2.2.2 Sample collection

In order to obtain optimal results, it is critically important that the collection of the sample is done with devices designed to:

- Be DNA-free and DNAse-free;
- Facilitate automation, preservation, and storage;
- Maintain chain of custody and sample integrity;
- Not interfere with the DNA analysis step;
- Achieve optimization in a variety of configurations for challenging forensic collections.

Numerous DNA sample collection kits are available on the market, aimed to address the above mentioned issues and to provide ready-to-use and easy-to-use solutions. Still, it is necessary to adopt specific precautions. To give few examples, a list of recommendations and good practice provisions is reported below:

- Particular precautions must be taken for fluids like semen, saliva and blood in the form of stains. Like on any "wet" sample, microbes are expected to grow soon on the swab, and thus facilitate DNA degradation. As a consequence, swabs should either be dried (avoiding contamination) or DNA should be extracted as soon as possible. In case of dry fluid stains, it is recommended to collect them by using a DNA-free swab moistened with a saline solution.
- For bones and teeth, it is recommended to use single use forceps to collect the sample and to place it in an appropriate envelope, as bones and teeth are particularly susceptible to contamination by handling. Moreover, because DNA is collected from tooth pulp and the solid bone, if possible, it is preferable to collect the whole tooth/bone.
- Hair should be plucked and not cut as most of high-quality DNA is present in hair follicles. It is thus recommended not to touch the hair follicle and, if possible to collect a minimum of 20 hairs from each individual.
- For tissue samples that include muscle or organs, it is recommended to collect a reasonable amount (about 1 cm^3). Muscle is preferred but any tissue (except gut because of the presence of gut microbes) is a good source of DNA.

Labelling is another relevant step to take into account and that might affect DNA profiling. Once the sample is secured, it is important to:

- Ensure that all sample envelopes are clearly labelled with the sample type, the collection site (GPS details if possible), date and name of collection officer, together with a full subsequent chain of custody.
- Provide accompanying paperwork which should be completed and enclosed with the samples.
- Ensure that samples and paperwork are sealed in a labelled, tamper-evident bag before transport.
- In case of shipping of samples, the packaged samples should be sealed and signed by the collector. The seals should be examined following transportation to ensure the package was intact during shipment. It is a good practice to sign samples submitted to the laboratory for analysis as evidence and placed in a secure evidence room.

Last but not least, in case of a concomitant legal context, it is recommended to ensure that duplicate samples are taken.

2.3 Step 2: From sample to profile

2.3.1 The DNA extraction step

To perform a forensic DNA analysis, DNA needs to be extracted from a sample. Usually half a nanogram of DNA is enough to provide good data. In general, it is desirable to have an extraction methodology that enables:

- The isolation of DNA from biological samples that contain small quantities of biological material;
- Obtaining the DNA at a high concentration so that the volume of extract used for subsequent procedures (see Section 2.3.2) is minimal;
- The removal of inhibitors or substances that interfere with the subsequent procedures;
- The extraction of DNA from a variety of biological samples;
- The adaptation of the manual protocol and its chemistry to automation.

An optimal procedure of DNA extraction should be non-toxic, fast, economic, and should enable to recover high purified DNA from each sample. However, there is no "universal" DNA extraction procedure, which fulfils all these requirements. Like for the sample collection step, in recent years various commercial kits for extraction of forensic DNA samples have been developed and are available on the market.

More recently, there is the possibility of automating the DNA purification process by using bench-top automated systems that enable the isolation of DNA from sample lysates.

These systems are very efficient also for the extraction of DNA from skeletal remains, which are considered among the most challenging ones. Moreover, the use of automated DNA extraction allows the standardization of tests and results in forensic laboratories, increasing the throughput and minimizing the risk of sample mixing due to human error.

2.3.2 The DNA quantification step

Once DNA is extracted, its quantification (i.e. determination of its amount) is important in forensic DNA analysis. The amount of DNA isolated by the extraction step needs to be quantified to ensure that the optimal amount of DNA is used for the next steps. Estimation of DNA concentration is particularly valuable, as STR amplification kits (used in the next step) are designed to work with specific DNA template ranges for optimal profile generation. Quantification of DNA by spectrophotometry or fluorimetry is insufficiently precise for forensic purposes, therefore PCR quantification method should be used (bearing in mind the need to move DNA quantification step from pre-PCR to PCR area of the laboratory).

Several quantification kits are available on the market. For forensics use, the kits should optimally:

- Contain pre-formulated internal amplification control (IAC) for each reaction, to allow quick monitoring of amplification success and to identify those samples that may contain inhibitors;
- Be compatible from the chemistry point of view with commonly used extraction technologies;
- Have been validated for human identification applications;
- Allow to quantify DNA from a wide variety of sources;
- Quantify a broad range of DNA concentrations (range 0.01 ng/µL ->10 ng/µL).

2.3.3 The DNA amplification step

Developed by Kary Mullis in 1983, the polymerase chain reaction (PCR) is still today considered a valuable tool in DNA forensics. Briefly, PCR replicates in millions of copies specific nucleotide sequences (demarcated from left to right by starters of PCR called "primers") from low amounts of DNA. Products of amplification ("amplicons") are then separated using electrophoresis according to their size. The amplification process continues to be used in more advanced techniques, including the massively parallel sequencing of whole genomes.

In forensics, the PCR technique is used to amplify the regions representing the STR markers and electrophoresis is then used to separate the DNA fragments. As explained in Section 1, the STR markers used in human identification have high variability among individuals and are measured by the lengths of the different alleles. PCR of STRs also allows for "multiplexing", which enables the analysis of several different loci at the same time, in the same tube.

Several commercial kits typically designed to amplify sets of STR loci that cover European and/or USA standard sets (Section 4.3), are available on the market; the latest ones routinely characterise more than 20 STR loci in a single experiment [40]. It has been observed that, sometimes, two different kits may give different results, i.e. the value of one allele may be different. In these cases, a third kit from a different supplier can be used to understand the correct value. This highlights the importance of recording information on the kit used for the DNA profiling, included as a field in some standard formats for DNA profiles (see Section 6.4).

In the past few years, instrumentations providing fully automated DNA profiling systems have been introduced. These instruments integrate the steps of DNA extraction, rapid PCR multiplex amplification (15 or more STR loci), amplicon separation, detection, sizing and genotyping in less than 90 min with, however, higher costs compared to conventional laboratory testing.

Fundamental intrinsic limits exist with PCR amplification due to random variation in sampling each allele at a locus. These stochastic events can lead to the so-called "allele drop-out" (i.e. failure to detect an allele). "Allele drop-in" (i.e. detection of a false positive allele) may also occur when the number of PCR amplification cycles is increased to improve sensitivity (see Section 6.3.2).

2.3.4 The use of massively parallel sequencing

Developments in the field of massively parallel sequencing, MPS (previously called next generation sequencing, NGS) have moved from original clinical genetics applications to the forensic field. If applied to missing person identification, the proper validation of the technique (see for example, [41]) and proper nomenclature [42] must be followed to also allow backward comparisons with data generated by the current STR detection standard (PCR and capillary electrophoresis).

2.4 Step 3: Evaluation of the genetic evidence

This final phase involves the comparison of the DNA profile attached to a missing person alert with the profile generated from a candidate missing person that needs to be confirmed and quantification of the evidence strength. Evidence strength is calculated using the **Likelihood Ratio** (LR), which represents the ratio of probabilities under two compared scenarios. When reference sample DNA profile is available for a missing person then there are two probabilities to be combined: a probability of obtained DNA evidence given that two samples belong to the same person, which is divided by a probability of obtained DNA evidence given that two samples belong to different persons. When sample DNA is available from volunteer family members instead of missing person themselves, scenarios and conditional evidence probabilities are modified accordingly.

In all cases, though, the evidence is evaluated by principles of inferential logic, underpinned by theoretical and empirical research in population genetics. This allows the probability of obtained DNA evidence given that two samples belong to different persons to be estimated from frequency of DNA sequences from these particular regions of the genome to be estimated with confidence for various populations. The allele frequencies of the different markers vary within the different populations, and the appropriate database should be chosen for calculation of denominator of likelihood ratio [43]. The correct population reflects the pool from which the unknown person is taken if they are not our missing person.

2.4.1 Direct comparison of two profiles

If the DNA markers of two compared profiles have different values for at least three loci, this is normally sufficient to exclude the possibility that the two DNA samples derived from the same person. If the marker values are identical, in forensics statistical analysis it is necessary to quantify the strength of evidence, to determine what is the ratio of the probability that the match is a consequence of the examined sequence coming *from the same individual* who provided the original sample and the probability that the match is a consequence from a randomly occurring *two individuals* in the general population.

Likelihood ratios that are currently achievable using the current kits are frequently in the order of billions allowing identity to be assigned with considerable confidence in many cases. In the simplest case, the LR is inversely proportional to the genetic profile frequency (LR = 1 / frequency of DNA profile in the population of potential profile donors). As observed by Professor B.S. Weir (North Carolina State University): "*In forensics, there is a clear distinction between "identity", meaning a unique existence, and "individualisation", pointing to a specific person. The forensic question is neither "Is this profile unique?" nor "Are these two profiles identical?", but "Is there sufficient evidence to demonstrate that these two profiles originate from the same source?*""[4].

[4] Genomes, Editors: Gustafson, J. Perry, Flavell, R.B. (Eds.), ISBN 978-1-4615-4235-3.

The minimal number of STR markers needed to achieve sufficient levels of confidence derives from circumstances of the case (from the priors), and from both theoretical and empirical studies. For example, a study of a profile dataset assembled by the Australian forensic agencies, which allowed more than 100 million comparisons of pairs of profiles, resulted in a single instance of eight loci matching between two persons, a father and son. All of the instances of nine loci matching revealed that they either came from identical twins or the same person profiled by different agencies [44]. A study of expected match probability using 13 loci (those from the original CODIS set) showed that the chance of generating random matches between unrelated persons in a database of 100 million different profiles is 10^{-15} [45].

These numbers are generally reflected in the size of the established Standard Sets (Section 4.3), explaining also the recent increase from 7 to 12 markers in the ESS. Under Prüm, a profile must include at least 6 of the ESS markers, to ensure that the comparisons can be performed with at least 6 common markers with the receiving country. Although this has caused issues of false-positives as the databases grew (see Section 5.2.1), for a missing person scenario according to the use of the SIS, this is mitigated by the appropriate estimation of "prior odds", i.e. the probability of identity based on non-genetic evidence [46] that would allow to calculate posterior odds by incorporating genetic evidence in the form of the likelihood ratio..

2.4.2 Kinship analysis

When comparing the profile of a missing person candidate with reference profiles provided in the alert by family members, the statistical analysis is more complex. Indeed, the profiles in this case are expected to be different: a child inherits half of their autosomal markers from each parent, distributed randomly (See Figure 2B). It is for this scenario that makers in standard sets are selected based on their low mutation rates (Section 4.3) to ensure that no changes occur to the individual marker values in the time frame when the genetic information is passed between the generations compared.

In this case, the LR is calculated based on a ratio of the probabilities of the DNA evidence given two hypotheses: 1) the candidate person is a member of the pedigree defined by the reference profiles and 2) the candidate person is unrelated to the known reference members of the pedigree (see [47]). These calculations can be very complex [48].

When using the DNA profiles of family members to help in the identification of missing persons, the best is to use as many family members as possible; unfortunately this can make the process costly. For any tangible case, it is possible to calculate potential LR in case of non-exclusion for different genotyped members of the pedigree using the appropriate software (see [49]). To provide some guidance, there have been studies to determine what are the best relatives to be chosen for the identification of missing persons (when available).

Starting from LR calculations, 37 common references scenarios have been analysed using the Caucasian population data on the 13 CODIS STR loci [47] and it emerged that the most likely combinations of reference relatives giving on average higher LR values are, in order (starting with the best):

1. Parents (possibly both - if both parents are typed, no other relative should be necessary);
2. Children (if the missing person is a male, sons are better because the Y chromosome is shared between father and sons);
3. A child and the biological father or mother of the child (who could be, for example, the spouse of the missing person);
4. Full siblings (if the missing person is a male, brothers are better because both the Y chromosome and mtDNA are shared, see Figure 3 and Figure 4);

5. Aunts, uncles, half-siblings, grandparents and cousins (less informative but they can be useful if one considers their Y chromosomes and mtDNA to increase the LR or to exclude false relationships).

In all cases, it is better if the DNA compared with that of the missing person is provided by at least two relatives whenever possible. It was also noticed that having reference profiles from volunteers that are relatives of the missing person but that are not related to each other gives higher average LR values.

The minimal number of markers to be used in these analyses is also greater than needed to match two profiles, usually in the range of 10-15 autosomal STRs [40]. For example, profiles may be submitted to the International Commission on Missing Persons (ICMP, see Section 4.2.2) database only if they comprise a minimum of 11 loci, plus Amelogenin [12]. Most modern kits allow the simultaneous analysis of at least 20 DNA markers, so for family members (i.e. with the availability of a good source of biological material), these numbers should be expected in the DNA profiles.

It is also in this context that lineage markers, such as Y-STRs and mtDNA, can be useful to provide further evidence (lineage LR) to autosomal LR values. However, care must be taken during combination of autosomal and allosomal LRs: lineage marker is not informative when the alternative hypothesis can concern lineage relative of the missing person.

Section 2. Summary of key concepts

- In general, best practices for the DNA profiles encourage the use of the highest number of markers possible (e.g. at least 20 autosomal STRs).

- If a good reference profile can be generated from the missing person, other reference DNA profiles (i.e. from kin) are not necessary and should not be included.

- For kin, the order of preference is: parents, children, siblings. Further relatives (e.g. grandparents, aunts, cousins, etc.) could still be acceptable if it is justified by the fact that no better options are available.

- When a living person is available (e.g. family members), DNA profiles should derive from either blood samples or buccal swabs as they are the best sources of quality DNA.

- As much technical information as possible on the process that generate the profiles, including kit and system used, should be included as they are useful information for the authority using this information.

- If permissible under the rules of the SIS, the population background of the missing person and their kin should be included to allow proper statistical analyses when matching profiles.

3 International Standards for DNA profiling methodologies

As the previous Sections illustrated, the DNA markers so far used in the establishment of DNA profiles constitute a minute fraction of the total information contained in the genome of each individual - at best, a few hundred bases out of 9 billion. Importantly, it is very hard and impractical to conclude about whether two DNA profiles correspond to the same (or related) person if they do not contain information about the same DNA markers. For this reason - and others related to the technical challenges involved in the generation of the profiles (see Section 2) - international working groups and organisations have been mandated to provide harmonised standards and quality criteria in the field. Some of the most relevant ones are presented in this Section.

3.1 International Criminal Police Organization-INTERPOL

INTERPOL is the biggest international police organization in the world, currently including 194 countries. The official name, since 1956, is the International Criminal Police Organization (ICPO–INTERPOL).

INTERPOL was officially created in Vienna in 1923 by Dr Johannes Schober, with the name of International Criminal Police Commission, and underwent various changes through the years. In 1989 the headquarters was moved to Lyon, France. In 2002, the I-24/7 web-based communication system launched[5], through which the member countries can access INTERPOL's automated DNA database.

After the 2004 Tsunami in Thailand, an INTERPOL project (FASTID) was started, creating an international 'Missing Persons/Unidentified Dead Bodies' database. INTERPOL has tested the use of the Bonaparte software [50] to compare the DNA profiles of missing persons with those of their relatives.

The official website of INTERPOL is www.interpol.int

3.2 The Federal Bureau of Investigation

The DNA Advisory Board (DAB) was a group originated under the DNA Identification Act of 1994 in United States of America by the director of the Federal Bureau of Investigation (FBI). It operated for five years from 1995 to 2000 and dealt with matters which concerned forensic DNA applications in particular developing initial Quality Assurance Standards (QAS) used in the U.S. for the forensics DNA community.

The first meeting was on 12 May 1995, and the DAB members included forensics scientists, molecular geneticists, a representative from the National Institute of Standards and Technology (NIST) and a judge [51]. The aim of DAB was to give common guidelines in order to guarantee standards for quality assurance in DNA test and forensic data which were gathered. In 2000 the DAB's responsibilities ended and the Scientific Working Group of DNA Analysis Methods (SWGDAM) (see Section 3.2.1) became responsible for offering recommendations on revisions to the QAS as needed [51].

The DNA Identification Act[6] financed forensic laboratories to improve the quality assurance standards on DNA analysis and allowed the FBI to found the Combined DNA Index System (CODIS) which is a set of databases that help matching the different information on violent crimes[7].

The website for CODIS is www.fbi.gov/services/laboratory/biometric-analysis/codis

(5) https://www.interpol.int/About-INTERPOL/History
(6) https://laws-lois.justice.gc.ca/eng/acts/d-3.8/
(7) https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/july2000/quality-assurance-standards-for-forensic-dna-testing-laboratories

3.2.1 The Scientific Working Group of DNA Analysis Methods

The Scientific Working Group of DNA Analysis Methods (SWGDAM) is a team composed of 50 scientists who represent many scientific laboratories in the USA. It was established by the FBI in 1988 with the name of Technical Working Group on DNA Analysis Methods (TWGDAM). Establishing quality assurance guidelines in order to improve and standardise DNA analysis was the first assignment of TWGDAM.

In 1999 the name of Technical Working Group on DNA Analysis Methods (TWGDAM) was changed into Scientific Working Group of DNA Analysis Methods (SWGDAM). This organization aims to provide new perspectives in the field, to organize meetings with experts who focus on topic such as mitochondrial DNA, population genetics, statistics, and Y-STRs. Meetings take place twice a year.

The SWGDAM's responsibilities include suggestions on the revision of the Quality Assurance Standards for Forensic DNA Testing Laboratories and the Quality Assurance Standards for DNA Databasing Laboratories.

The official website of SWGDAM is www.swgdam.org

3.3 The International Organization for Standardization

The main standards related to the process of producing the DNA profiles include:

ISO 18385:2016 - *Minimizing the risk of human DNA contamination in products used to collect, store and analyse biological material for forensic purposes – Requirements*. As the name implies, this standard specifies requirements for the production of consumables and reagents to be used in the collection, storage, and analysis steps for the biological material used to generate DNA profiles. The aim is to avoid nuclear DNA contamination of these products when used by the forensics community, as described in Section 2.

ISO/IEC 17025:2017 - *General requirements for the competence of testing and calibration laboratories* - This standard is based on the ISO 9001:2015 (a standard that expresses requirements for an organization's quality management system), and is the minimum recommended accreditation for laboratories producing DNA profiles for a DNA database [40] (see Section 6.1)

ISO 21043:2018 - This series of standards deals with requirements for the different steps of the forensic process, starting from recognition and collection at the scene of crime to reporting the results in the courtroom. The first two have been published so far: *Part 1 – Terms and definitions* and *Part 2: Recognition, recording, collecting, transport and storage of items,* that cover the early steps of the process (Section 2.2). Future work items include *Part 3 – Analysis and Examination*, *Part 4 – Interpretation* and *Part 5 – Reporting*.

A standard is also available for the storage and exchange of the DNA profiles:

ISO/IEC 19794-14:2013 - *Information technology - Biometric data interchange formats - Part 14:DNA data*, which is described in more details in Section 6.4.3.

3.4 International expert groups

3.4.1 The European Network of Forensic Science Institutes

The European Network of Forensic Science Institutes (ENFSI) is an international group of experts in forensic science. The aim of the network is to share knowledge, exchange experiences and promote common standards in the field of forensic science in order to allow comparison of forensic research and investigation results between European countries.

The main activities of ENFSI are to organise meetings and scientific workshops, to contribute to other studies, to provide expertise on forensic issues and to publish best

practice handbooks. At the moment, ENFSI has 69 members in 37 countries, and hosts 17 Expert Working Groups including the Expert DNA Working Group.

The cooperation between European DNA laboratories has started before the birth of ENFSI. In 1988, the European DNA Profiling (EDNAP) was born in London, when a group of forensic scientists decided to join information and compare results about DNA analysis, in order to improve crime investigation. The ENFSI Expert DNA Working Group was established in Birmingham in 1995 and is collaborating with EDNAP since 2004.

The objectives of the ENFSI Expert DNA Working Group, which can be found on their official website[8], are:

- To bring together recognized (i.e. ISO 17025/IEC accredited organizations in accordance with Council Framework Decision 2009/905/JHA) organizations actively pursuing forensic DNA analysis methods for the purpose of exchanging and disseminating information on forensic applications;
- To discuss, share, and compare forensic DNA analytical methods, protocols and research;
- To establish quality assurance guidelines and quality controls for European forensic DNA analysis;
- To co-operate with other recognized national and international organizations in developing European standards for forensic DNA analyses;
- To serve as a mechanism for the review and revision of European guidelines for forensic DNA analyses;
- To disseminate to the European forensic DNA community ENFSI guidelines, forensic research results and any other work of benefit to the European forensic DNA community.

The official website of ENFSI is www.enfsi.eu

3.4.2 The International Society for Forensic Genetics

The International Society for Forensic Genetics (ISFG) is an internationally non-profit scientific organization established in 1968 in Mainz, Germany, by a group of expert who studied blood antibodies. At the beginning, this society took the name of International Society for Forensic Haemogenetics (ISFH). In 1999 the society was renamed ISFG because genetic methods became more widespread.

The ISFG aim is to diffuse knowledge about genetic markers used in the field of forensic sciences. This is done through the organization of biannual international congress, seminaries/workshops, and external quality controls at regional and international levels and publications in their journal *Forensic Science International: Genetics*.

The International Society for Forensic Genetics has published many guidelines pertaining to scientific standards on genetic laboratories techniques and in particular on genetic marker used in juridical systems.

The ISFG includes more than 1,200 members spread in over 60 countries.

The official website of ISFG is https://www.isfg.org/.

The ISFG DNA Commission was established in 1987 to deal with matters related to the use of DNA in forensic sciences. Its aim is to discuss the developments of DNA polymorphisms in relation to their use in the scientific and legal contexts and to make the appropriate recommendations. The creation of the Commission was related to massive parallel sequencing of forensic STRs. Among the currently active projects is "Evaluation of evidence (beyond the DNA profile)" that aims to relate DNA profile to 'how', 'why' or 'when' it becomes evidence.

The official website of the ISFG DNA Commission is

(8) http://enfsi.eu/about-enfsi/structure/working-groups/dna

www.isfg.org/Working+Groups/DNA+Commission

Section 3. Summary of key concepts

- The working groups and organisations described in this Section have produced a wealth of documents that summarise their work in setting guidelines and in providing recommendations in various aspects of the use of DNA profiles for forensics. Many of these have a direct relevance for this document. Annex I lists some of these important documents, as a suggestion for additional reading.

- The relevant conclusions will be further discussed and summarised in the following Sections.

4 Current practices in established data repositories of DNA profiles

While Section 3 focused on the platforms through which best practices in the field are being discussed and published, the aim of this Section is to provide an overview of some DNA profile databases currently operating in Europe, in non-criminal contexts, and the nature of the information they contain.

4.1 Differences between criminal and non-criminal DNA profile databases

A criminal DNA database contains DNA profiles deriving from crime scenes, convicted offenders, suspects and crime-stains, with the objective to solve a crime by matching crime-related stains to persons. In certain cases, a criminal DNA database might also contain DNA profiles of volunteers who have provided their DNA in order to facilitate investigations. Some countries allow the inclusion of deceased victim's DNA from unsolved cases in their databases.

A non-criminal database (such as for missing persons) contains DNA profiles of the missing persons, their relatives and unidentified human remains, with the objective to identify the missing. A second objective is to link different parts of the body between them as well as to the missing, e.g. in natural disaster situations, war crimes, and terroristic acts. Unidentified persons are not apparent victims and are usually included in missing persons' databases; nonetheless for identification purposes, these profiles might also be compared with criminal DNA profiles, in an attempt to identify the victims.

Because of the different objectives, some Member States keep the criminal and the missing persons databases separated, while others have one single integrated database [40]. Keeping these two databases separated present the following advantages:

- Data protection: the DNA profiles of missing persons and their relatives cannot be erroneously compared with the ones stored in a criminal DNA database;
- The two DNA databases can be managed by different organizations reflecting the different objectives;
- A specialised software is needed to find and evaluate matches between potential missing persons/unidentified human remains and the relatives of the missing person, which is different from the computing strength of evidence that two DNA profiles are from the same person (see Section 2.4).

It may still be useful, in cases where the DNA databases are separated, to compare the DNA profiles of unidentified human remains with the DNA profiles of the criminal DNA database, as:

- DNA profiles of unidentified human remains found in one location may match with stains found at a crime scene at another location, indicating a crime and transportation of the unidentified to another location;
- DNA profiles of unidentified human remains may match with a reference sample, which may assist identification. This comparison needs to be done only once, as the unidentified person is dead and hence cannot be added to the DNA database as a reference sample in the future.

4.2 Established missing persons databases

4.2.1 EU Member States

Most Member States have tried to produce legislation to regulate national forensic databases with distinct orientations on the establishment of criteria for inclusion and retention of profiles [52]. In some Member States, criminal databases and missing person's databases are merged whereas in others, these are very distinct with different criteria for inclusion and maintenance of a profile in the database [40]. While the Prüm

Treaty [53] is applied to the fight of crime and terrorism (see Section 5), in the case of missing persons, the situation is more complex and the criteria are not yet universal.

The inclusion of a DNA profile is dependent on the situation of the missing person; it could be due to mass fatalities from natural or human-induced causes e.g. a tsunami or a terrorist attack. Situations can be closed, where the numbers and relationships between the missing persons are known (e.g. plane crash), or open, where the number of missing persons cannot be assessed, e.g. tsunami. Samples derived from personal items of missing persons or samples obtained from family members are named *ante mortem* samples, while samples from unidentified bodies (or body parts) are *post mortem* samples [40].

The deletion criteria for database data vary across countries; in general there is maximum duration that the DNA profiles can be retained. It is important to take into consideration the fact that deleting a DNA profile might also require the destruction of the material from which the DNA was extracted, meaning that this profile cannot be generated again. The laboratory data, i.e. electropherogram might also have to be deleted together with associated documentation.

In summary, the criteria used by countries for the storage of a DNA profile are not uniform and are based on many parameters: e.g. time after inclusion, type of crime, repeated convictions. The storage time may last till the death of a person, fixed time after the death of a person, variable time after the death of a person depending on crime, fixed/variable time after the completion of a sentence depending on the crime, or until no longer relevant.

The ENFSI recommends that if the removal of a DNA profile from the DNA database is dependent on external information, a process should be in place to provide the custodian of the DNA database access to this information, preferably by means of an automated message, delivered after any event that influences the deletion date of a DNA profile [40].

4.2.2 The International Commission on Missing Persons

The International Commission on Missing Persons (ICMP) was initially established to help with the 40 000 missing persons resulting from the conflict in and breakup of Yugoslavia. The tasks of the ICMP were formulated with different mandates and, more recently, the ICMP treaty signed by 10 countries resulted in the establishment of the DNA laboratory in The Hague (January 2018), rather than in Sarajevo where it was previously located. In 2000, the ICMP started its attempts of DNA identification in a large scale; high-throughput autosomal STR testing from skeletal remains were conducted successfully [54].

DNA testing from degraded skeletal remains was primarily making use of mtDNA testing due to its high copy number and easiness of amplification; in the case of ICMP, though, there was no option to obtain direct *ante mortem* (AM) reference samples for the missing (biopsy samples, known personal effects, see Section 2.2.1), so the identification was performed by genetic kinship analysis of family members. Although mtDNA in its traditional usage offers advantages in that the distant maternal relatives can be used as references (Figure 4), given the scale of the event and the lack of distinctive non-DNA identification evidence in almost all cases, the resolving power of mtDNA would have been insufficient for the task of the ICMP. It was therefore decided that nuclear autosomal DNA testing was the only route (Section 1.3.1).

Three major points were required: 1) DNA extraction and amplification methods that would provide a sufficient success rate on degraded skeletal material, 2) DNA reference samples from multiple close references would have to be obtained for each missing person, and 3) the development of an effective software for large-scale kinship matching [12].

After the 2004 South East Asian Tsunami, in conjunction with INTERPOL, ICMP deployed its forensic expertise as part of a major international Disaster Victim Identification (DVI) program [55], creating a Forensic Data Management System (fDMS) that contains forensic data including data on missing persons and their relatives.

4.2.3 The Spanish Phoenix program

The "Phoenix Program" was launched in Spain in 1998 [56], with the objective to identify human remains from missing persons. It contains two independent databases: the database with the STR and mtDNA profiles from bones, and the reference database, which contains the STR profiles and mtDNA from relatives [30]. At the date of the publication, more than 3,700 families had contacted Phoenix, 862 had enrolled in the program and at least 319 unidentified remains had been identified and returned to the relatives. In order to obtain uniform and reliable results, the authors suggested recommendations for these types of databases:

- Standard operating procedures and universally accepted genetic markers should be used;
- For results to be reliable, laboratories should be subjected to quality assurance and quality control programs;
- The technology should be automated if possible in order to facilitate the typing of large volumes of samples and to permit national and international searches and comparisons;
- Proper use of the database to guarantee confidentiality according to national laws, informed consent from voluntary donors and court orders to handle human remains are some of the requirements of the Spanish database management.

4.2.4 The Committee on Missing persons in Cyprus

Cyprus' Committee on Missing Persons (CMP) was established in April 1981; the Greek Cypriot and Turkish Cypriot communities, under the auspices of the United Nations, agreed on its establishment and, since 1997, the leaders of the two communities agreed to provide each other all information at their disposal on the location of graves of Greek Cypriot and Turkish Cypriot missing persons.

In 2006, the CMP began excavations and exhumations on both sides of the island. In order to provide the required expertise, archaeologists and anthropologists from the Argentine Forensic Anthropology Team (EAAF) were brought in to coordinate and train a bi-communal team of Cypriot scientists involved in exhumations and anthropological analyses. An anthropological laboratory was setup in the United Nations Protected Area in Nicosia.

Since 2008, the CMP’s forensic team has been carrying out exhumations autonomously. EAAF forensic experts continue to be involved in the project for quality control purposes. Out of 1510 Greek Cypriots and 492 Turkish Cypriots missing, 681 and 246 missing individuals respectively have been identified and the remains returned to their families. The ICMP has also provided support and knowledge to the initiative; in July 2012 the ICMP began providing assistance in making DNA-based identifications. The ICMP assisted with DNA isolation from post-mortem samples received from the CMP and matched these DNA profiles against DNA profiles from anonymized family reference samples. In addition, the ICMP has provided guidance on problematic cases, and assistance in matching profiles from a historical database of samples produced prior to the ICMP’s involvement.

As part of this project, for both the generation of family reference samples and the population genetic studies, 18 autosomal STR loci and 17 Y-STR loci have been selected for the analyses [57].

4.2.5 Missing Persons Bureau-UK

The UK has established the Missing Persons Bureau (MPB) with the objective to identify missing persons and to provide adequate information to the relatives. The MPB database consists of a collection of DNA profiles from missing persons, their close relatives and unidentified people or human remains. It is separate from the National DNA Database.

DNA profiles are only used to identify people or human remains, not for other purposes (e.g. criminal investigations). Once a missing person's location is known or a person is identified, the DNA profiles are deleted. In case of no match, the person is not found and the DNA profiles are kept in the database for comparison against individuals or remains found in the future.

Any new profiles for unidentified individuals or remains submitted to the database in the future will be compared against the profile of the missing person already submitted. The profiles will be retained until the missing person is located or if consent to store the DNA profile is withdrawn[9].

4.2.6 The United Nations human rights and the Red Cross suggestions

In February 1980, the Commission on Human Rights decided to "*establish for a period of one year a working group consisting of five of its members, to serve as experts in their individual capacities, to examine questions relevant to enforced or involuntary disappearances of persons*". This mandate has since been renewed. The tasks of the Working Group on Enforced or Involuntary Disappearances are various; concerns/recommendations on the search of persons and their DNA have been expressed in the recent document "***Draft guiding principles for the search for disappeared persons***", where it is suggested that "*Procedures for the gathering of DNA samples require the prior and informed consent of the potential donors of the samples and the confidentiality of the victims, and ensure that the samples will be used exclusively to identify and locate the disappeared person*" [58].

Similar concerns have also been expressed by the International Committee of the Red Cross in its document "MISSING PERSONS: A Handbook for Parliamentarians" [59]; it is clearly stated that there must be a regulatory framework to ensure protection of sensitive information like DNA and privacy of both of the missing persons and their relatives. Moreover, national legislation should provide, in the cases where DNA samples are taken, the method for doing so and the processing of the data in the framework of the intended purpose. "*It is important to ensure that a DNA analysis performed for the purpose of identification of a missing person be separated from any other use, for example, criminal proceedings; otherwise it may inhibit recourse to this form of information gathering on the part of relatives and interested parties*" [59].

4.2.7 DNA-Prokids

The term missing children includes many categories of child disappearances such as runaways (children who leave their homes voluntarily), parental abductions, missing children in migration, criminal abductions, children trafficking, lost, injured or otherwise missing (children disappeared for no apparent reason) [60].

DNA-Prokids is an international project on the prevention of, and fight against, human trafficking using genetic identification of victims and their relatives. It was created in 2009 and currently its headquarters is at the University of Granada, in Spain[10].

DNA-Prokids is divided in three tiers [30]:

1. A national level with two genetic databases, one for the DNA profiles (and metadata) obtained from children who live in an illegal situation and the other for DNA profiles (and metadata) voluntarily provided by relatives.

(9) https://www.missingpersons.police.uk/en-gb/resources/factsheets-for-families
(10) http://www.dna-prokids.org/

2. An international level with a database that allows sharing genetic data among countries.
3. The third level is focused on data generation. Data on children found outside of their family, children who are going to be adopted and data of voluntary relatives of missing children will be automatically inserted into a database for comparison.

Scientific technology tools to find missing children are available, but often the legal and social (privacy) difficulties hinder this process. For this reason it is important to try to implement the coordination and standards (legal and scientific) among the states. Moreover, cooperation and communication among countries can deter criminals from kidnapping children [30].

4.3 Standard Sets of Short Tandem Repeats

The established DNA profile databases described above operate by generating profiles using specific sets of genetic markers. As explained earlier, it is meaningless to compare two or more DNA profiles unless they contain information about the same DNA markers.

The institutions and working groups described in Section 3 have produced different standard sets of DNA markers, described below, with their overlap shown in Table 2. The number of markers present in these sets is a trade-off between being small enough (to be analysed in a single experiment) versus being sufficient (to provide sufficient discriminating power). There are many common considerations in the selection of what constitutes a "good" marker to be inserted in a standard set, which include [45]:

- No known association with medical conditions or defects (for privacy reasons);
- Low mutation rate (probability less than 0.30%);
- Absence of linkage and linkage disequilibrium (high level of independence, i.e. having one of the markers does not increase the probability of having one of the others);
- High level of discrimination (a locus with a probability of identity preferably of less than 0.1);
- It must already be in use by the international forensic DNA community;
- Successfully tested in inter-laboratory exercises;
- Compliance with quality assurance standards (refers to the loci satisfying the requirements of the FBI Director's Quality Assurance Standards such as validation, being human specific, etc.).

4.3.1 The CODIS Core Loci

The CODIS Core Loci is a set of STR loci to be used in the FBI system of the same name (see Section 3.2). Before 2017 the minimum set of loci required by CODIS, called the "CODIS core", consisted in 13 loci (CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51 and D21S11) and Amelogenin. It was recently expanded, by the CODIS Core Loci Working Group, to include seven additional loci (D1S1656, D2S441, D2S1338, D10S1248, D12S391, D19S433 and D22S1045) [61].

4.3.2 The European Standard Set (ESS)

A 1998 INTERPOL initiative promoted the use of DNA information to register sexual offenders (in particular child molesters), with a core of 4 agreed loci: TH01, vWA, FGA and D21S11; these became known as the European standard set of loci (ESS). One year later, the set of STR loci for the INTERPOL register and the ESS was expanded to include three more, namely D3S1358, D8S1179 and D18S51 [62]. These loci, confirmed in 2001, constitute the core of all national DNA databases in Europe [63]. Since then, the ESS of STR loci has been increased to 12 in 2009, by adding D1S1656, D2S441, D10S1248, D12S391 and D22S1045 [64].

4.3.3 The INTERPOL Standard Set

Because of their close development, the INTERPOL Standard Set of Loci (ISSOL) has been the same as the ESS, with the addition of Amelogenin (Section 1.3.3). The five additional loci added to the ESS were also included to the ISSOL in 2010 [65]. In addition to these 13 loci, the INTERPOL search request form allows 16 additional loci, not technically part of the ISSOL, shown in a separate column of Table 2.

Table 2. Summary of the STR loci present in the different published standard sets, including the additional loci included in INTERPOL's DNA profile information form

LOCI	STANDARD SETS			ADDITIONAL LOCI (INTERPOL DNA profile information form)
	CODIS	ESS	ISSOL	
CSF1PO	X			X
FGA	X	X	X	
TH01	X	X	X	
TPOX	X			X
vWA	X	X	X	
D3S1358	X	X	X	
D5S818	X			X
D7S820	X			X
D8S1179	X	X	X	
D13S317	X			X
D16S539	X			X
D18S51	X	X	X	
D21S11	X	X	X	
D1S1656	X	X	X	
D2S441	X	X	X	
D2S1338	X			X
D10S1248	X	X	X	
D12S391	X	X	X	
D19S433	X			X
D22S1045	X	X	X	
AMEL	X		X	X
Penta E				X
Penta D				X
SE33	::[1]			X
FES				X
F13A1				X
F13B				X
CD4				X
GABA				X

[1] *Although the SE33 marker it is not officially part of the CODIS Core Loci, the CODIS system, starting from version 7.0, allows information for this marker to be used (i.e. inserted and searched for) in DNA profiles.*

Source: JRC analysis, 2019

4.4 Establishing family links in reference samples

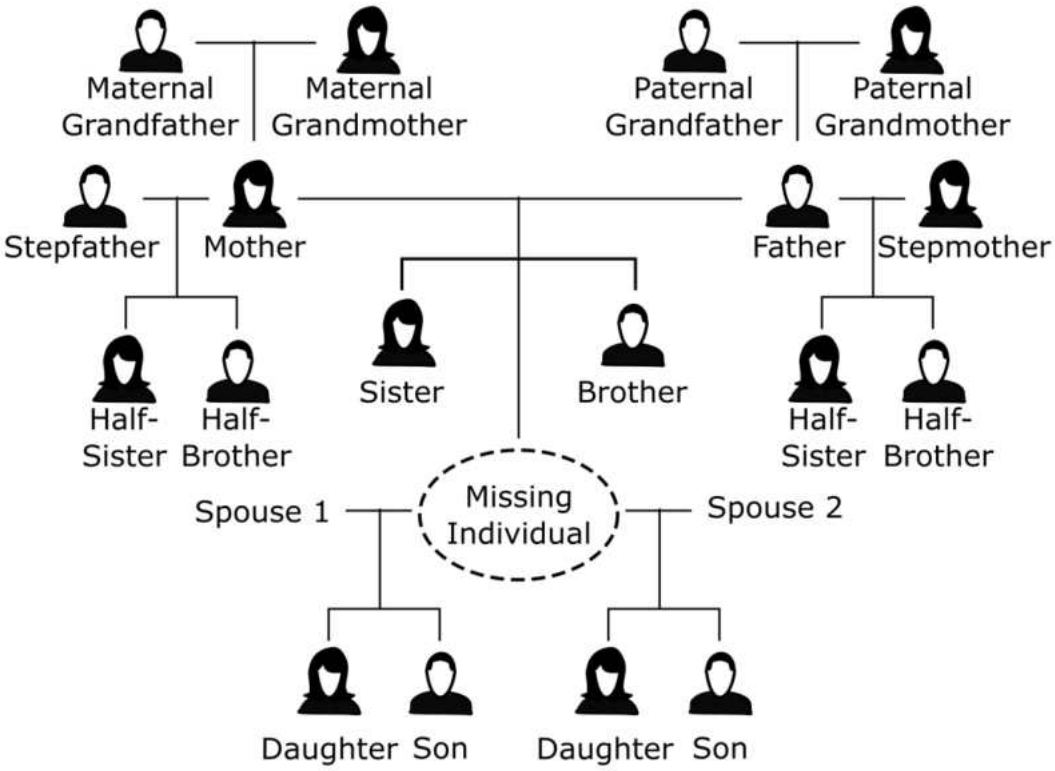
It is important for any initiative establishing and maintaining a DNA profiles database for missing persons to correctly understand the victim's family structure when using DNA to establish biological relationships and identify the missing persons. The complexity of properly capturing the exact links between the different family members, and between the family members and the missing person, should not be underestimated.

Different factors were reported contributing to this complexity [66,67], including:

- The language and terms used to describe biological relationships can be confusing and may vary according to the cultural context (e.g. "cousin"). This has led to efforts, for example in Electronic Health Records, to harmonise the vocabularies for family relationships (see, for example, SNOMED CT[11] and HL7's RoleCode[12])
- Family members are under extreme stress when relatives go missing, potentially causing them to provide incorrect or inaccurate information;
- The description of family links can become difficult when the donors of reference profiles are distant relatives;
- Complicated family structures exist, involving, for example, multiple marriages, adoptions, *in vitro* fertilizations, etc.

For this reason, when the DNA profiles of family members are gathered for the identification of missing persons, it is recommended and common practice to report, not only a written description (father, uncle, etc.) but a family tree where the biological link between the donor and the missing is made explicit (see an example in Figure 5).

Figure 5. Example family tree as part of a DNA reference profile collection form. The donor is invited to circle their relationship to the missing individual.



Source: JRC, 2019, adapted from [67]

[11] https://browser.ihtsdotools.org/
[12] https://www.hl7.org/fhir/v3/RoleCode/cs.html

Section 4. Summary of key concepts

- Criminal and non-criminal DNA profile databases differ in the type of information they content and their purposes.
- The institutions and working groups described in Section 3 have produced different standard sets of DNA markers, according to specific selection criteria.
- The number of markers in these sets is a trade-off between being small enough to be analysed in a single experiment and large enough to provide sufficient discriminating power.
- Established DNA profile databases operate by generating and storing profiles using specific sets of genetic markers, as it is difficult to compare two or more DNA profiles unless they contain information about the same DNA markers.
- If family members reference profiles are used, it is crucial to correctly understand the victim's family structure when using DNA to establish biological relationships and identify the missing persons. A family tree showing the biological links between the missing person and the reference profile donor(s) should be included as additional information.

5 Lessons learned from the exchange of DNA profiles under Prüm

The previous Section described some of the existing databases containing DNA profiles in Europe. A current, practical example of systematic exchange of DNA profiles between EU Member States can be found in the activities around the implementation of the Prüm decisions. This Section summarises the challenges that have been (or are being) faced in the implementation of this framework, highlighting when this experience could lead to recommendations for the SIS.

It should be noted that DNA profiles of missing persons and relatives are not exchanged under the Prüm. For this exchange, other police cooperation channels are usually used.

5.1 The Prüm Decisions

In May 2005, Austria, Belgium, France, Germany, Luxembourg, the Netherlands, and Spain signed the Convention on the stepping up of cross-border cooperation, also known as the "Prüm Convention" [68]. Other Member States joined the first group of signing countries and, presently, there are 14 States that have ratified the convention and 5 countries that have expressed their intention to do so [69]. In June 2008, some parts of the Prüm Convention became the Council Decision 2008/615/JHA [70], with Council Decision 2008/616/JHA [53] defining its implementation. These are collectively referred to as the "Prüm Decisions".

According to Prüm Decision 2008/615/JHA, each Member State must elect a National Contact Point (NCP) which has the task to check the exchange of data with the other States [71]. In order to prevent crimes and to maintain public order, Member States exchange data during major events (such as cross-border events or regarding a possible terrorism offence).

The system of DNA data exchange does not have a central site; the databases, software application, and e-mail components are located within the Member States and the data exchanged on a peer-to-peer basis.

The flow of data during a request under the Prüm Decisions [72] begins when a Member State generates a message with a DNA profile, encrypts this message through Secure/Multipurpose Internet Mail Extensions (sMIME) before sending it by e-mail to the requested country. The network system that connects all Member States is called sTESTA (Trans European Services for Telematics between Administrations). After receiving the requested message, the requested country decrypts the message and searches its national database with the attached profile. The answer consists of "HIT" or "NO-HIT" result (assuming no error occur in the process); this answer is what is sent back to the requesting country, through the same encryption steps and network. Even in the case of a HIT notification, the response does not contain any personal information - further investigations will be carried out by an identification number that allows the exchange of more detailed information on the results by specially authorized officers.

5.2 Challenges in the implementation of the Prüm Decisions

According to the original decisions, all EU Member States should have implemented the Prüm Decisions by August 2011. However, a 2018 study highlighted that some Member States were still in the process of implementing them [71]. The reasons for these delays were analysed in different documents (see, for example, the Council of the European Union Note 14918/10 [73]). The main challenges met are: IT problems, privacy and data protection issues, legal issues, national structures, lack of information, lack of human resources and funding [69]. Some of the most relevant ones are further described below.

5.2.1 Challenges related to the exchange of DNA profiles

According to the State of Play report (5017/3/16) of the Council [74], ten countries were reported, at that time, to have problems in actively and successfully exchanging DNA data [69]. Technical and non-technical reasons can be associated to the problems in exchanging DNA data.

Number of overlapping loci

As described in Section 4.3, three standard sets of loci are used by the different countries to generate DNA profiles:

- A set of 20 loci in the Combined DNA Index System (CODIS);
- A set of 12 loci in the European Standard Set (ESS);
- A set of 24 loci in the INTERPOL DNA Profile information form (ISSOL + additional).

With the advent of the Prüm Decisions, and the consequent numerous exchanges of DNA profiles, the lack of uniformity among the standard sets used by different Member States in their national databases has become an issue. In fact, only 12 loci (those used in the ESS) are common to all standard sets (see Table 2); the EU Council resolutions 2001/C 187/01 and 2009/C 296/01 encouraged Member States to use the ESS as a minimum in their national databases, in order to allow comparison of DNA profiles.

Impact for the SIS: The DNA profiles uploaded in the CS-SIS, no matter from which country they are submitted should contain as many loci as possible from the 12 common loci included in the European Standard Set.

False positives/negatives

According to the Prüm inclusion and matching rules [53], it is necessary that at least six loci match between both DNA-profiles before a "Hit" response is provided by the requested authority. However, the automatic comparison of DNA profiling increased the risk to find false-positive matches because of a significant growth of DNA profiles in national DNA databases. Some have suggested considering using more loci (8 to 10) to decrease false positives.

Impact for the SIS: Based on this experience, a recommended number of 10 loci should be requested for the minimum number of loci to be included in a DNA profile for the CS-SIS. This is **not** applicable when using reference DNA profiles (i.e. family members of the missing), as the statistics are completely different (Section 2.4) and, in this scenario, the minimum of loci necessary should be higher.

5.2.2 Challenges related to the follow-up of requests made under Prüm

After the comparison of the DNA profiles, the hits obtained are subject to different evaluation criteria (not only by strength of evidence by likelihood ratio but also by tactical, reliability, legal, priority and sustainability issues); as a consequence, Taverne et al.(2017) have stated that only 2% of the total number of "Hit" responses following a request under Prüm were used in court [75].

Impact for the SIS: Although not directly applicable for the SIS, this highlights that sometimes it is difficult to reach a conclusion with sufficient confidence when comparing DNA profiles to ascertain identity. This is even more delicate when family member profiles are used. Often, it is necessary to use, thus to generate and share, additional information (more markers, more references, etc.) to what is originally attached to a request. In some cases, allowing the authority that uses the DNA profiles in a missing person case to contact the authority that generated it can be crucial for a final conclusion to be drawn.

5.3 Future steps: the European Forensic Scientific Area

On 15 June 2011, at the meeting of the Law Enforcement Working Party in Brussels, Pawel Rybicki, Chairman of ENFSI, presented the proposal to create by 2020 a European

Forensic Science Area This proposal was endorsed by the EU Council [76]. The European Forensic Science Area will be an infrastructure where routine forensic processes will be based on harmonised scientific and legal standards and in which forensics experts will cooperate with each other and with the criminal justice systems.

Member States and the Commission will work together to make progress in different areas [77], including the establishment of common best practice manuals and their application in daily work of forensic laboratories and institutes and the identification of optimal and shared ways to create, update, and use forensic databases.

The ENFSI and Europol will participate to the initiative to establish the European Forensic Science Area; the Cross-Border Exchange and Comparison of Forensic DNA Data in the Context of the Prüm Decision (DAPIX) will take care of activity 6: "Stimulating exchange of forensic data via Prüm and improving its quality" [71].

Section 5. Summary of key concepts

- The Prüm Decisions describe a communication system to allow the exchange of DNA profiles between the different Member States and their individual DNA profile databases.
- DNA profiles of missing persons and relatives are not exchanged under the Prüm. For this exchange, other police cooperation channels are usually used.
- The specific STR markers to be used for the alerts in the CS-SIS should be a combination of those present in the Standard Sets, to maximise compatibility between authorities.
- The DNA profile, if generated for the missing person, should be composed of at least 10 autosomal STR markers.
- The information and contact details of the laboratory that generated the profile should be attached to the submitted profiles, in case further information is needed when matching the profiles to a missing person candidate.

6 Evaluating the quality of a submitted DNA profile

Once generated through the process described in Section 2, a DNA profile is represented by alphanumeric data consisting of a set of markers names, each associated to a pair of numbers (the number of repeats at each of the two alleles of the marker). For mtDNA, the profile is a string of letters representing the sequence of the DNA at the region used for forensics. Figure 6 shows an example of a DNA profile, both with the STR markers (A) and a mtDNA sequence (B).

Figure 6. Example of a DNA profile as generated for forensics purposes, including (A) the repeat values of different STR markers and (B) mtDNA sequence of the HV1 region

A

Marker	Allele 1	Allele 2
vWA	17	18
TH0I	6	9
D21S11	29	31
FGA	23	26
D8S1179	14	15
D3S1358	15	15

Marker	Allele 1	Allele 2
D18S51	14	18
D13S317	11	12
D7S820	8	11
D5S818	11	11
D10S1248	10	18
Amelogenin	X	Y

B

>NC_012920.1:16024-16569
TTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACAT
TACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATC
AAAACCCCCTCCCCATGCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCC
ACCCCTCACCCACTAGGATACCAACAAACCTACCCACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATA
GCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGGTCCCTTGACCACCATCCTCCGT
GAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAACTGT
ATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG

Source: JRC, 2019. (B) taken from the reference mitochondrial genome, GenBank Accession NC_012920).

A DNA profile thus consists of text, which needs to be taken into account when identifying the specific metrics/criteria that can be developed to ensure that "*the minimum data quality standards and technical specifications have been met",* as stated in Regulation (EU) 2018/1862 [3]. This Section describes a set of strategies that have and can be used to ensure the quality of the DNA profiles to be stored in DNA databases.

6.1 Demonstrating the quality of the process that generated the profiles

Sections 2.2 and 2.3 highlighted how complex and delicate the whole chain of events involved in the production of DNA profiles from the biological sources is. Mistakes made at any point of the process may result in a final profile which is either incorrect or incomplete. This is reflected in the intense and coordinated efforts, through the years, to develop sets of standards, best practices and quality management rules in the field and in the laboratory (Sections 2 and 3).

Although the DNA profile database manager, who receives the final product, generally has little control or oversight of what is done upstream, it is possible to use these standards as a way to document (and verify) the quality context of the process.

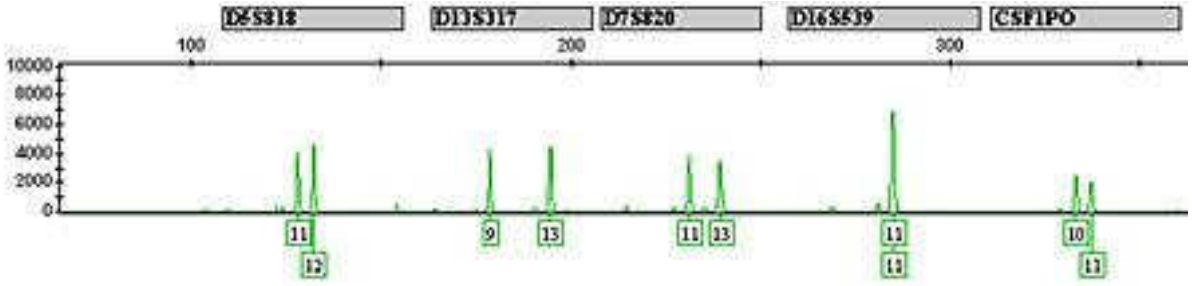
As described in Section 3.3, ISO/IEC 17025:2017 allows laboratories that carry out tests, sampling, and calibration of the instruments to demonstrate that their results are reliable and valid; this allows the sharing of results among laboratories in different countries, without having to carry out further tests. The standard covers General Requirements, Structural Requirements, Resource Requirements, Process Requirements, and Management System Requirements [78]. Quality assurance and control measures consist of several elements, including documentation and validation of methodologies, internal and external proficiency testing, and periodic case review. Laboratories can demonstrate that they are adhering to international standards through third-party accreditation.

ISO/IEC 17025:2017 is the minimum recommended accreditation for laboratories producing DNA profiles for a DNA database [40].

6.2 Verifying the quality of the experiment that generated the profile

Although the final "processed" DNA profile consists of alphanumeric data (see Figure 6), the values are generated by the interpretation of the output of a laboratory instrument. Commonly, the value (number of repeats) for each STR allele is determined by comparing the size of the fragments to a standard size ladder and allelic cocktail, resolved by electrophoresis. An example output trace is shown in Figure 7.

Figure 7. Example electropherogram of an experiment to generate a DNA profile, which is interpreted from comparing the size of the alleles at the different markers with standard size



ladders and allelic cocktail.

Source: Kline/NIST, taken from https://www.nist.gov/image/cellularfingerprintjpg

There is, at this stage, a lot of information and automatic software available for the experienced operator to interpret and take into account when transforming these traces to the final numbers composing the DNA profile, which include indications regarding the quality of the experiment.

Tools have also been developed to specifically evaluate the raw electrophoresis data contained in the files generated by the instruments at this stage, such as the OSIRIS (Open Source Independent Review and Interpretation System) tool[13]. The type of quality control metrics generated by this tool from an experiment includes the measured noise for each channel, the max base pair error (from sample to ladder), and the peak heights [79], providing an independent assessment of the experiment quality.

For sequencing experiments (for example in reading the mtDNA sequence), similar tools have been developed and are available, depending on the sequencing technology used.

Although quantified quality metrics can be objectively evaluated to a set of thresholds/ranges of acceptance, from the point of view of a DNA profile database manager the following concerns should be kept in mind:

- Even with the assistance of the software tools, this evaluation requires highly specialised and advanced technical skills in the field;

([13]) https://www.ncbi.nlm.nih.gov/projects/SNP/osiris/

- The tools/metrics/thresholds need to be adapted to the different "raw" outputs produced by different (existing and future) technologies used in the process;
- The raw data files are usually large in size and require an appropriate storage infrastructure;
- The files cannot be used as such to compare different DNA profiles in confirming the identity of a missing person. An interpretation to a marker/value text format is in any case necessary for this.

For these reasons, this approach is not normally considered in DNA profile database management procedures. However, a DNA database manager should not give up regarding the incoming data available, see below.

6.3 Checking for "hints" of low quality in the final text file

Although the final DNA profile is "simply" text, there are a lot of opportunities in the values it contains to evaluate its quality and hint at some problems that may have appeared upstream. This Section summarises some of the strategies evaluating the values present in the final profile.

6.3.1 Contamination

Because the most common procedure to derive the values of the different markers (STRs or mtDNA) involves an amplification step that starts with a low amount of the donor DNA, any contamination of the original sample with DNA or biological material of someone involved in the process prior to the amplification step may interfere with the final profile. This is the reason why avoiding contamination is so important in the process (see Section 2.1.2). If this happens despite the precautions taken, two things may happen:

1. The final profile is a combination of the "correct" profile and the contaminating source;
2. The contaminating source completely takes over and the final profile corresponds to a single, but wrong, person.

Both these scenarios may be tested in the final DNA profile, as explained in the following paragraphs.

Profiles representing a combination different sources: mixed profiles

The term "Mixed profile" refers to the presence of DNA information from two or more individuals in a single DNA profile [80]. Scientifically this phenomenon is observed when DNA analysis by capillary electrophoresis shows an electropherogram (Figure 7) with three or more peaks at more than two loci. If the sample is originated from only one individual (barring very rare biological phenomena) the peaks are at most two, one for each allele [81].

In order to standardise as much as possible the interpretation of these profiles in the forensic field, guidelines have been defined. When mixed profiles of the victim and perpetrator are seen from sexual assault evidence, the victim profile can be generated separately and subtracted from the mixture [82]. In other cases, however, even using complex statistical approaches that calculate likelihood, mixed profiles interpretation remains a complex process that may not generate usable evidence [83,84]. Best practices documents state that if a mixed profile cannot be avoided in a DNA profile database, it should not contain more than 2 individuals [40]. Under the rules of the international exchanges under Prüm, no mixed profiles are allowed [85].

Profile corresponding to the wrong person: elimination databases

Elimination (or exclusion) databases are DNA profile databases containing the profiles of the actors involved in generating a particular DNA profile, from the crime scene workers to the laboratory staff. If a match is found between the generated profile and someone in

this database, it can be inferred that this profile was derived from a contamination along the process and is thus not representative of the original biological sample.

A 2011 initiative by the *Laboratoire de Sciences Judiciaires et de Médecine Légale* (LSJML), in Québec (Canada), aimed to increase the number of crime scene workers volunteering their own DNA profile for an elimination database [86]. The analyses showed that 14% of the profiles uploaded in the resulting elimination database could be matched to profiles generated in previous crime scenes. This highlights the importance of this procedure, now included in the best practices in many Member States. As an example, Ireland's "Criminal Justice (Forensic Evidence and DNA Database System) Act 2014" describes in part 5, the procedure in taking samples to generate elimination databases from both the Garda Síochána and crime scene investigators[14].

Best practices also recommend the inclusion in elimination databases of forensics laboratory staff of all categories, as far as possible within the boundaries of the data protection law of their country [40]. Even the workers in the chains of production of consumables and kits used by the forensics laboratories may be a source of contamination. The ICMP has been involved in a series of discussions with the relevant companies in order to develop a Manufacturers' Exclusion Database (MED) to be used in this context [87].

6.3.2 Low quantity or poor quality of the purified DNA used to generate the profile: dropped alleles.

Allelic drop-out and drop-in events are forensic phenomena that occur when DNA extracted from a sample and used to generate a profile is in very low amount or strongly degraded.

Allelic drop-out happens when the system manages to amplify just one of the two alleles in a heterozygous genotype. The result of this is an incorrect characterization of that given locus that will contain only one value instead of two; in this case, in fact, a heterozygous genotype will be genotyped as homozygous. For example, a specific marker that should have the values 10 and 15 would be reported as having the values 10 and 10, if the conclusion of seeing a single peak (of value 10) is that both alleles are merged in the same peak. Locus drop-out happens when the system completely fails to amplify the given locus, usually locus with long amplicons.

Drop-in is another irreproducible artefact that consists in the synthesis, during the PCR reaction, of one or more alleles different from those actually present. These result in more than 2 values for a specific marker (or more independent markers), which sometimes may be mistaken with the presence of a mixed profile [88].

Despite numerous studies [88–90] that suggest guidelines and software that allow probabilistic analyses (e.g. EuroForMix, STRmix, TrueAllele, and Kongoh for quantitative continuous model, LRmixStudio for qualitative model), to date, there is no universally accepted interpretation of these artefacts in the forensic field.

6.3.3 Consistency of the marker values with population information

Although, when a sufficient number of markers is characterised, the resulting profile can be assigned to a unique person, the values for individual markers fit within relatively narrow ranges representing the existing variation at these positions of the genome that exist in the human species.

As more and more DNA profiles are being generated, either as part of forensics cases or scientific studies, better information becomes available about the range of possible values for each marker and their relative frequencies in different populations. Databases are being maintained to make these results available, as they are important in the computations involved in comparing the DNA profiles, and include pop.STR [91], STRidER

(14) http://www.irishstatutebook.ie/eli/2014/act/11/enacted/en/html

[92], STRBase [93] and ALFRED [22]. Table 3 shows, when available, the range of values reported in the different resources for each of the markers in Table 2. Although rare allele values outside these ranges may occur, this information can be used to at least highlight possible abnormal values in a DNA profile that would warrant a second look at the raw data files.

Table 3. Range of reported values for each allele in databases compiling information about the frequencies in different populations. The last column shows the upper and lower limits for each marker that can be used to compare the values in a final DNA profile

LOCI	ALLELE FREQUENCIES				Recommended verification range
	pop.STR	STRIDER	STRBase	ALFRED	
CSF1PO	6-15	/	5-16	5-18	5-18
FGA	16-46.2	16-33.2	12.2-51.2	9-45.2	9-51
TH01	5-12	5-10.3	3-14	4-31.2	3-32
TPOX	6-13	/	4-16	5-16.1	4-17
vWA	11-22	11-21	10-25	9-24	9-25
D3S1358	6-20	11-20	8-20	7-44.2	6-45
D5S818	7-16	/	6-18	6-22	6-22
D7S820	6-15	/	5-16	5-19	5-19
D8S1179	8-18	8-18	7-20	6-25	6-25
D13S317	7-16	/	5-17	5-18	5-18
D16S539	5-15	8-15	4-16	4-16.2	4-17
D18S51	9-28	9-25	7-39.2	5-31.2	5-40
D21S11	24.2-39	24.2-35.2	12-41.2	12-41.2	12-42
D1S1656	8-20.3	9-20.3	9-21	9-19.3	8-21
D2S441	8-17	8-17	8-17	/	8-17
D2S1338	11-27	14-28	11-28	10-29	10-29
D10S1248	8-19	9-19	8-19	/	8-19
D12S391	12-27.2	14-27	15-26'	13-28	12-28
D19S433	9-18.2	10-18.2	5.2-20	7-20	5-20
D22S1045	8-19	10-20	8-20	/	8-20
Penta E	5-24	/	5-26	5-29	5-29
Penta D	2.2-17	/	1.1-18	3.2-22	1-22
SE33	3-34.2	6.3-32.2	3-39.2	6.3-37	3-40
FES	7-15	/	7-15	4-16	4-16
F13A1	3.2-17	/	3-17	1-19	1-19
F13B	3-12	/	6-12	5-12	3-12
CD4	/	/	4-15	4-16	4-16
GABA	/	/	/	/	/

Source: JRC analysis, 2019, compiled from [91], [92], [93] and [22]

The same analyses can be performed for common Y-STR markers (Table 4).

Table 4. Range of reported values for commonly used Y-STR alleles from the STRidER database

MARKER NAME	ALLELE RANGE* (REPEAT NUMBERS)
DYS19	10-19
DYS385 a/b	7-28
DYS389 I DYS389 II	DYS389I: 9-17 DYS389II:24-34
DYS390	17-28
DYS391	6-14
DYS392	6-17
DYS393	9-17
YCAII a/b	11-25
DYS388	10-18
DYS426	10-12
DYS434	9-12
DYS437	13-17
DYS438	6-14
DYS439	9-14
DYS447	22-29
DYS448	20-26
DYS456	13-18
DYS461 (A7.2)	8-14
DYS635 (C4)	17-27
Y-GATA-H4	8-13 (25-30)
Y-GATA-C4	20-25
Y-GATA-A10	13-18

Source: JRC analysis, 2019, compiled from [92]

6.3.4 Mitochondrial DNA sequence: the FASTQ format

In the case where the "value" of a marker is not a number but a DNA sequence, as in the case of the hypervariable regions of mitochondrial DNA (see Figure 6B), a standard file format, the FASTQ format, allows sharing both the sequence and an associated "quality score" for each base in that sequence, the FASTQ format [94]. A sequence in FASTQ format is made by four lines per sequence, for example:

```
@SEQ_ID description
CTCGCATCATCAGCTAGCATCGATCATCGATCAGTCACGTAGTC
+
!''*((((***+))%%%++)(%%%%).1***-+*''))**55C!
```

where:

- Line 1 begins with a '@' character and is followed by the sequence identifier and an optional description.
- Line 2 is the raw sequence letters, i.e." A","T","C" and "G".
- Line 3 begins with a '+' character and is optionally followed by the same sequence identifier (and any description) again.
- Line 4 encodes the quality values for the sequence in Line 2, and must contain the same number of symbols as letters in the sequence of line 2. To each base in the string describing the DNA sequence (A, C, T or G) corresponds a byte representing the quality of the sequencing result at that base, also described as a number or a corresponding ASCII character: from 0x21 (lowest quality; '!' in ASCII) to 0x7e (highest quality; '~' in ASCII).

Several bioinformatics tools to analyse FASTQ files and produce reports on quality of sequences are available. One of the most used is FASTQC, freely available at https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

It should be noted that, for convenient reporting of mtDNA haplotypes and for database searches, a format has been proposed that only reports the differences in the DNA sequence relative to the revised version of the first human mtDNA sequence (rCRS)[15]. When transforming the mtDNA sequence to this format, it is important to follow the proper convention and rules [95]. However, since this format does not include indication of the quality of the original sequencing experiment, and can be, in any case, derived by the downstream user of this information from the full sequence, it is recommended to request the full sequence (in the FASTQ format) for inclusion in the CS-SIS.

6.4 Standardised formats for DNA profiles

The fact that both the data and the metadata associated to DNA profiles can be expressed as alphanumeric text allows the use of mark-up languages as an efficient format to store and share DNA profiles, and automated rules can thus be incorporated to verify the consistency of the information within the different tags.

In fact, different XML standard schemas have been described for the purpose of exchanging DNA profiles information, summarised in this Section. The development of these formats are crucial to allow the development of automatic export/import procedures, as manually entering DNA profiles in databases has been shown to be a significant source of error [40].

6.4.1 Prüm

Council Decision 2008/616/JHA, that implements the "Prüm" legislation (see Section 5.1), includes in its Annex (Chapter 1, Section 4 of [53]) a description of an XML schema for the exchange of DNA profile information (Figure 8).

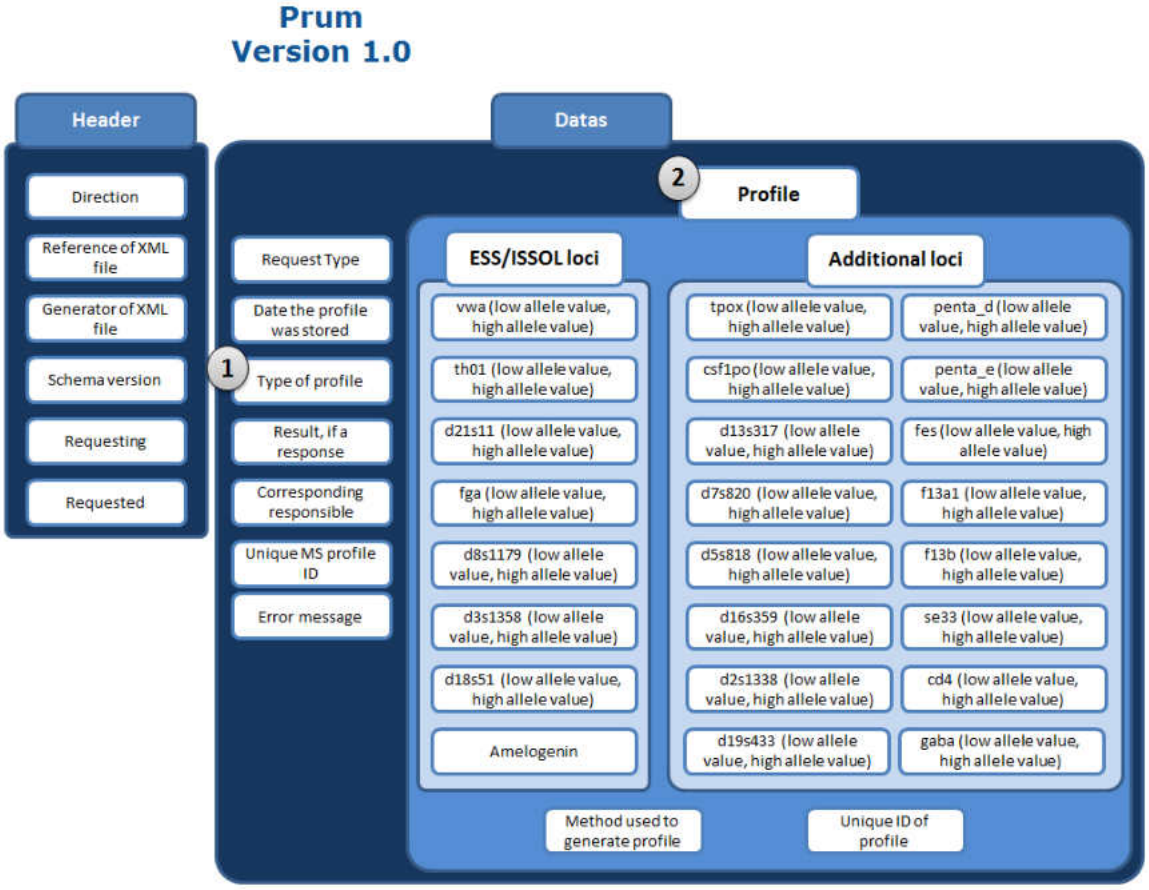
The header fields contain information about the transmission defined specifically for the Prüm DNA exchange (request, requesting and requested). The body ("datas") contains the DNA profile(s) relevant to the request, and is based on the schema defined for the INTERPOL DNA exchange gateway.

This schema is thus designed to send specific requests (or the response), focusing on the investigation of criminal offences. This has a few consequences for the use of attaching DNA profile information to assist in the identification of missing persons:

- Specimen info limited to "person" or "stain", with no possibility to include family relationships in case of reference profiles.
- The schema is limited to ESS/ISSOL loci, with a set list of possible additional loci, those listed in Table 2. These are named fields: the names of the 24 markers are coded as the names of the tags in the schema. For this reason, the format cannot accommodate other markers, including any Y-STRs.

[15] https://www.ncbi.nlm.nih.gov/nuccore/NC_012920

Figure 8. Schematic representation of the structure and fields of the Prüm XML format. Highlighted are (1) the field for the type of profile ("person" or "stain") and (2) the fields to include the information about the specified set of DNA markers.



Source: JRC analysis, 2019, from [53]

6.4.2 CODIS

The CODIS Unit of the Federal Bureau of Investigation has published two Interface Specifications for the exchange of DNA profiles within their systems. The first, the CODIS CMF 3.2[16], describes the interface between CODIS and external systems. The second, the recently published CODIS Rapid Import CMF[17] describes the interface between the Rapid DNA instruments and the CODIS Rapid Enrolment (CRE) application (Figure 9).

Both formats are similar and can accommodate any STR marker, autosomal or Y-STRs. They allow inclusion of information about the kits used in the laboratory and, in the case of the Rapid Import CMF, the instrument that produced the profile.

These formats have fields that contain information on the specimen related to the profile which is more descriptive than the Prüm format, and in the case of the CMF 3.2 specification, explicitly allows family links information (Biological Child, Biological Father, Biological Mother, Biological Sibling, etc., see appendix D of the specification file).

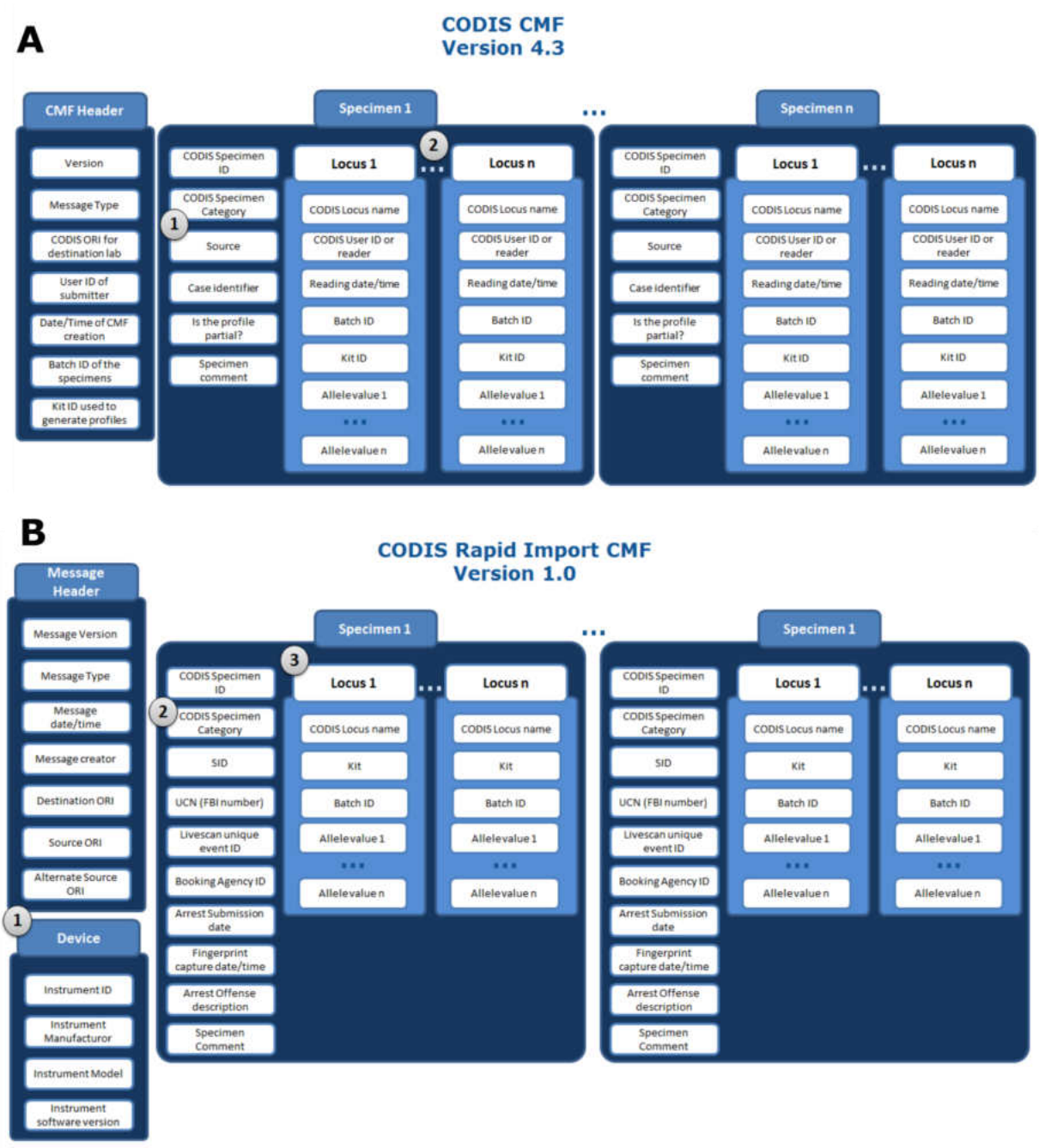
Because of the Rapid DNA Concept of Operation described in the specification, Rapid Import CMF includes required fields linked to arrest time and crime description, as well as

(16) Currently accessible from: https://github.com/ncbi/osiris/blob/master/osiris/docs/CODIS%20CMF%203.2%20Interface%20Specification.pdf

(17) http://www.fbi.gov/file-repository/codis-rapid-import-cmf-interface-specification-r16-170925-508.pdf

reference to fingerprint capture, which is not applicable for the CS-SIS under the current legislation. The explicit list of valid specimen categories is also thus more limited (Arrestee, Convicted Offender, Detainee, Juvenile, and Legal), but the appendix leaves the possibility of using other categories if needed.

Figure 9. Schematic representation of the structure and fields of the CODIS XML formats. A: the CODIS CMF version 4.3; highlighted are (1) the description of the source of the profile (specimen) and (2) the information about the markers analysed for each specimen. B: the CODIS Rapid Import CMF Version 1.0; highlighted are (1) the information on the device used to generate the profile, (2) the description of the source of the profile (specimen) and (3) the information about the markers analysed for each specimen.



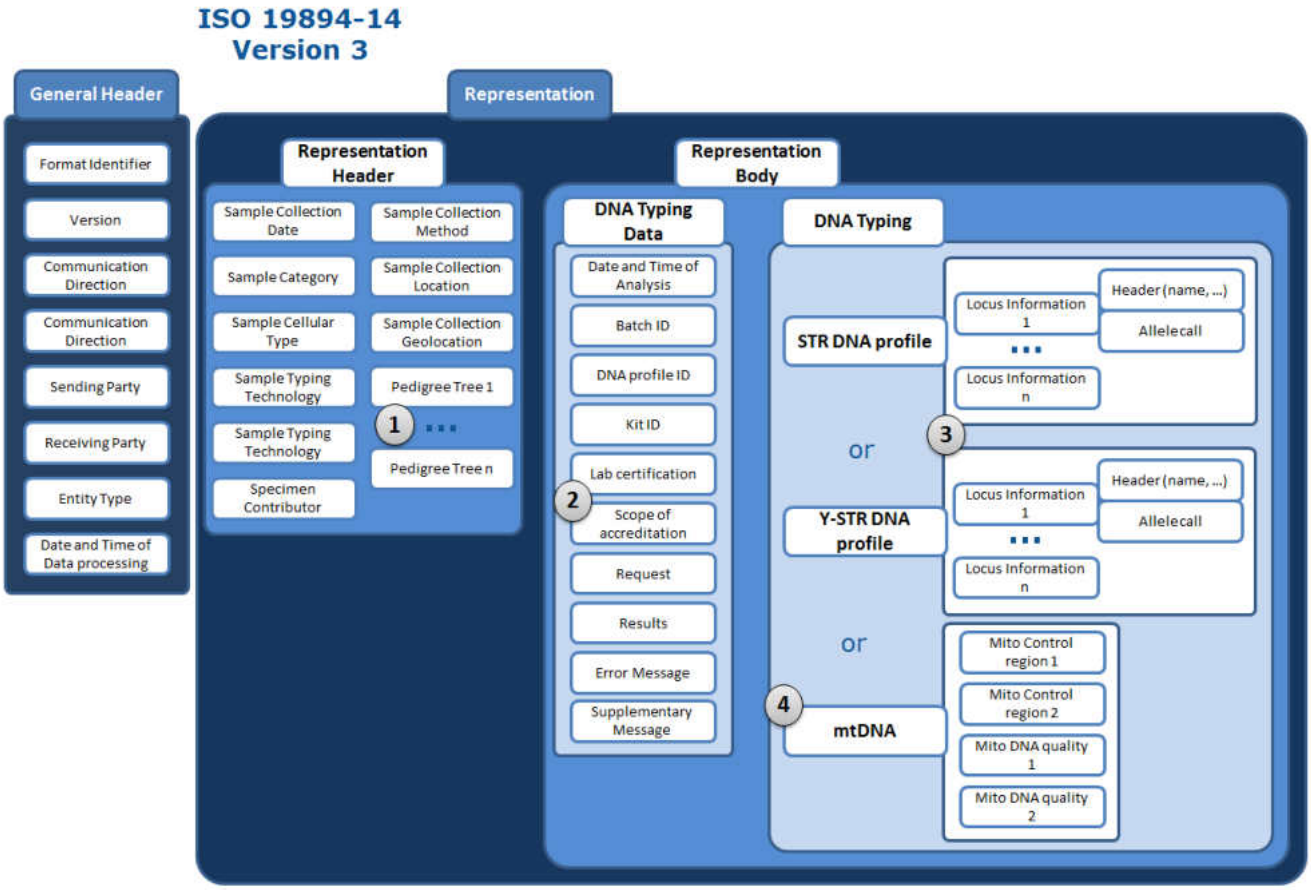
Source: JRC analysis, 2019

6.4.3 ISO/IEC 19794-14

The ISO/IEC 19794 series is a set of standards designed to describe biometric data interchange formats applicable to identity management systems. In this series, part 14 refers specifically to the exchange of DNA profiles for the purpose of person identification or verification [96].

This standard is the most flexible of those described in this Section, allowing not only the description of any STR marker, but also mtDNA sequences (Figure 10). In addition, it allows unequivocal description of the links between the missing person and the family reference samples (see Section 4.4), including information about family members who did not contribute a profile, by allowing the description of the full relevant pedigree. It also contains fields to describe the certification of the laboratories, including the scope of this certification (e.g. for STRs, mtDNA, or both).

Figure 10. Schematic representation of the structure and fields of the ISO 19794-14 XML format. Highlighted are (1) the full pedigree related to the profile(s) included, (2) the certification of the laboratory that generated the profile, (3) the information about the markers analysed for each sample, if they are STRs and (4) sequence information, if the marker used is mtDNA.



Source: JRC analysis, 2019, from [96]

6.4.4 Verifying internal consistency

An important advantage of using structured formats such as those described above is that rules can be implemented to automatically verify the integrity and completeness of the submitted DNA profiles. An amendment to the ISO/IEC 19794 Standard, released in 2016, added an Annex A that describes an exhaustive conformance testing methodology, that included rules to verify that the information in the different fields are in line with what is expected.

If, on the other hand, entry of a DNA profile in the CS-SIS is done manually, the process should have special functionalities in order to minimise the risk of a human error (mistyping). For entering profiles into CODIS and Interpol, this is done by asking to enter the whole DNA profile twice, and the profile is accepted only if the values of loci are the same in all fields in both attempts.

Other conventions can be implemented to further increase the chances of identifying errors in the process that generated the profile. One example is the common convention that the two values for each STR marker should be reported in the order of their value, i.e. first the smaller, then the larger of the two (e.g. "12, 15" and not "15, 12"). These rules can be added to those verifying the plausibility of the individual allele values (see Section 6.3.3).

Another possibility to verify internal consistency is the evaluation of multiple reference profiles in the case they are biologically related. As an example, if the reference profiles are said to include both the father (A) and the full-sibling sister (B) of the missing person, it is possible to check and confirm, by comparing their profiles, that A is the biological father of B.

Section 6. Summary of key concepts

- A DNA profile is represented by alphanumeric data. If STR markers are used, it consists in a set of markers names, each associated to a pair of numbers (the number of repeats at each of the two alleles of the marker). If the marker used is mtDNA, it consists of a DNA sequence.
- The DNA profiles to be stored in the CS-SIS should be expressed as alphanumeric text, formatted using a mark-up language such as XML.
- Different standard XML formats already exist, and may be used for this purpose. It is recommended to investigate with the different authorities the possibility for their DNA database software to export in one of these existing formats (in particular ISO 19794-14), to avoid manual entry of the information.
- If manual entry of a DNA profile is allowed, the process should have special functionalities in order to minimise the risk of a human error (mistyping). For example, the profile should be entered twice, and the values should be consistent. Also, for each STR locus the first reported allele value should be, as a rule, lower than the second allele value.
- Validation rules, mapping acceptable values for the different fields in the submitted profile (test assertions) should be developed and applied.
- The laboratories that produced the profile(s) must be ISO/IEC 17025:2017 (and/or nationally equivalent) accredited.
- No mixed profiles should be submitted (max 2 values for each STR marker), unless justification is provided.
- Because of the availability of living persons (thus high quality samples), no dropped alleles should be allowed for reference profiles from family members.

- The DNA profiles should be checked against appropriate elimination DNA databases.
- For markers submitted as DNA sequences (i.e. mtDNA), the format should include a quality value for each base (the FastQ format), that should allow conclusions of an acceptable quality level for forensics purposes.
- The names of the loci included in the profile should be checked against the list of standard markers. The values for each locus included should be verified against the range of known values for these profiles in the population, when available. Discrepancies should be investigated.
- When uploading multiple reference DNA profiles that have established familial links between each other the consistency of the kinship relationships between these reference individuals should be confirmed.

7 Conclusions and Recommendations

This Section summarises the recommendations and conclusions for the addition of DNA profiles to alerts submitted to the CS-SIS. The right column refers to the Section of the text where the statement is taken from.

These focus on the two main objectives of this report: 1) What is the nature and content of a "DNA profile" that may be attached to alerts on the CS-SIS (data, metadata and format); 2) How can their "quality" be evaluated?

7.1 DNA profiles - data

1	The DNA profiles attached to missing persons alerts in the CS-SIS should be composed of STRs, described as the name of the markers attached to their values (number of repeats on each of the two alleles).	*Section 1.3.1, Figure 6*
2	The use of other markers, such as SNPs, is not expected at this stage, due to the state of the art of the harmonised implementation of this technology.	*Section 1.3.5*
3	The specific STR markers to be used should be a combination of those present in the Standard Sets, to maximise compatibility between authorities.	*Section 4.3, Section 5.2.1*
4	The submissions should encourage the use of the highest number of markers possible (at least 20 autosomal STRs).	*Section 2.4*

If a good quality reference DNA profile is available directly from the missing person:

5	The profile should be composed exclusively of autosomal STRs.	*Section 1.3.1*
6	The profile should be composed of at least 10 autosomal STR markers.	*Section 5.2.1, Section 2.4.1*
7	Reference DNA profiles (from kin) are not necessary and should not be included.	*Section 2.4.1*

If a DNA profile is not available from the missing person, or if there are doubts about its origin or concerns about its quality:

8	Reference profiles, from available ascendants and/or descendants, should be included following the appropriate consent procedures.	*Section 4.2, Section 2.4.2*
9	If available, the order of preference should be: parents (especially mother), children, and siblings.	*Section 2.4.2*
10	Further relatives (e.g. grandparents, aunts, cousins, etc.) could still be acceptable if no closer kin options are available.	*Section 2.4.2*
11	Each profile should be composed of at least 20 autosomal STR markers.	*Section 2.4.2*
12	DNA profiles from family members should be derived from either	*Section 2.2.1*

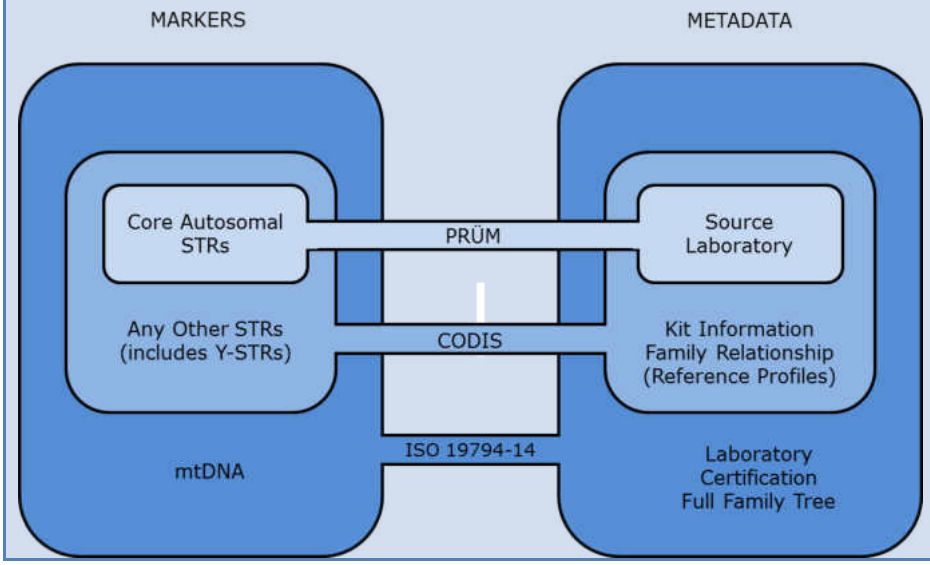
	blood samples or buccal swabs.	
13	Because of the possibility of directly sampling the persons (thus high quality samples), no mixed profiles or dropped alleles should be allowed for reference profiles.	*Section 6.3.1, Section 6.3.2*
14	For reference profiles, the use of other markers (Y or X chromosomes STRs, mtDNA) should be allowed, if they make sense based on their inheritance pattern relative to the missing person.	*Section 1.3.2, Section 1.3.4, Section 2.4.2*

7.2 DNA profiles - metadata

In addition to the DNA profiles per se (markers and values), the submission should include the following information:

15	The information and contact details of the laboratory that generated the profile, in case further information is needed when matching the profiles to a missing person candidate.	*Section 5.2.2*
16	The certification of the laboratory that generated the profile(s).	*Section 6.1*
17	If reference profiles are used, a family tree showing the biological links between the missing person and the reference profile donor(s).	*Section 4.4*
18	As much technical information as possible, including kit and system used to generate the profiles.	*Section 2.2, Section 2.3*
19	If permissible under the rules of the SIS, the population background of the missing person and their kin should be included to allow proper statistical analyses when matching profiles.	*Section 2.4*

7.3 DNA profiles - format

20	The DNA profiles should be expressed as alphanumeric text, formatted using a mark-up language such as XML.	*Section 6.4, Figures 9-10*
21		*Section 6.4, Figures 9-10*

	Different standard XML formats already exist, that may be used for this purpose. The figure above (source: JRC, 2019) summarises the relevant data/metadata they can describe.	
22	We recommend investigating with the different authorities the possibility for their DNA database software to export in the ISO 19794-14 format, to avoid manual entry of the information.	*Section 6.4*
23	If manual entry of a DNA profile is allowed, the process should have special functionalities in order to minimise the risk of a human error (mistyping). For example, the profile should be entered twice, and the values should be consistent.	*Section 6.4.4*

7.4 Quality of the submitted DNA profiles

Different options to verify the quality of the submitted profiles are available and summarised in Section 6. From these, the following are proposed for the SIS alerts, either to be performed by the submitter prior to submission (and **stated** in the submission), or upon reception in the CS-SIS (by the **CS-SIS** database manager).

Upstream process

24	The laboratories that produced the profile(s) must be ISO/IEC 17025:2017 (and/or nationally equivalent) accredited (**stated**).	*Section 6.1*
25	The laboratory that generated the profiles should maintain the original sample/purified DNA/raw output of experiment (in particular for a sample from the missing person), until the alert has been resolved, in case additional information/clarification is needed by the receiving laboratory when a match is attempted (**stated**).	*Section 5.2.2*

Contamination and experiment quality

26	No mixed profiles should be submitted (max 2 values for each STR marker), unless justification is provided (**CS-SIS**).	*Section 6.3.1*
27	The DNA profiles should be checked against appropriate elimination DNA databases (**stated**).	*Section 6.3.1*
28	For markers submitted as DNA sequences (i.e. mtDNA), the format should include a quality value for each base (the FastQ format), that should allow conclusions of an acceptable quality level for forensics purposes (**stated**).	*Section 6.3.4*

Consistency of data

29	Validation rules, mapping acceptable values for the different fields in the submitted profile (test assertions) should be developed and applied (**CS-SIS**).	*Section 6.4.4*
30	The names of the loci included in the profile should be checked against the list of standard markers (**CS-SIS**).	*Table 2*

31	The values for each loci included should be verified against the range of known values for these profiles in the population, when available. Discrepancies should be investigated (**CS-SIS**).	*Table 3, Table 4*
31	For each STR locus the first reported allele value should be lower than the second allele value (**CS-SIS**).	*Section 6.4.4*
32	When uploading multiple reference DNA profiles that have established familial links between each other the consistency of the kinship relationships between these reference individuals should be confirmed (**stated**).	*Section 6.4.4*

7.5 General comments

33	Considering the active development of standards and procedures in the field conducted by various international institutions and working groups, it is recommended to review the recommendations of this report related to standards on a regular basis.	*Section 3, Section 1.3.5*
34	There should be considerations to link the information with other established missing persons databases, as fragmentation of information is undesirable for the efficiency of missing persons identification.	*Section 4.2*

7.6 Future steps

35	The conclusions and recommendations of this report are focused on the use of DNA profiles to help the identification of missing persons. If the use of DNA profiles is ever extended to, for example, Article 40 (unknown wanted persons), it is recommended to independently review the state-of-the-art for this specific use case.	*Section 4.1*
36	In case the use of DNA profiles is, in a future revision of the legislation, extended to Article 40, we recommend the use of dedicated independent databases for the storage of DNA profiles in criminal (Article 40) and non-criminal (Article 32) use cases.	*Section 4.1*
37	The objective of the current study is related to the storage of DNA profiles, under the existing legislation. Should future revisions consider the implementation of a search engine for DNA profiles, we recommend that an additional study evaluating the readiness and availability of these technologies be performed.	*Section 2.4*

8 References

[1] Council of the European Union. Council Decision 2007/533/JHA of 12 June 2007 on the establishment, operation and use of the second generation Schengen Information System (SIS II) 2007.

[2] The European Parliament and the Council of the European Union. Regulation (EC) No 1987/2006 of the European Parliament and the Council of 20 December 2006 on the establishment, operation and use of the second generation Schengen Information System 2006.

[3] The European Parliament and the Council of the European Union. Regulation (EU) 2018/1862 of the European Parliament and the Council of 28 November 2018 on the establishment, operation and use of the Schengen Information System (SIS) in the field of police cooperation and judicial cooperation in criminal matters, amending and repealing Council Decision 2007/533/JHA, and repealing Regulation (EC) No 1986/2006 of the European Parliament and of the Council and Commission Decision 2010/261/EU 2018.

[4] Lo Y-T, Shaw P-C. DNA-based techniques for authentication of processed food and food supplements. Food Chemistry 2018;240:767–74. doi:10.1016/j.foodchem.2017.08.022.

[5] Loreille O, Ratnayake S, Bazinet A, Stockwell T, Sommer D, Rohland N, et al. Biological Sexing of a 4000-Year-Old Egyptian Mummy Head to Assess the Potential of Nuclear DNA Recovery from the Most Damaged and Limited Forensic Specimens. Genes 2018;9:135. doi:10.3390/genes9030135.

[6] Bartlett JMS, Stirling D. A short history of the polymerase chain reaction. Methods Mol Biol 2003;226:3–6. doi:10.1385/1-59259-384-4:3.

[7] Pang AW, MacDonald JR, Pinto D, Wei J, Rafiq MA, Conrad DF, et al. Towards a comprehensive structural variation map of an individual human genome. Genome Biology 2010;11:R52. doi:10.1186/gb-2010-11-5-r52.

[8] Straub RE, Speer MC, Luo Y, Rojas K, Overhauser J, Ott J, et al. A Microsatellite Genetic Linkage Map of Human Chromosome 18. Genomics 1993;15:48–56. doi:10.1006/geno.1993.1008.

[9] Butler JM. Genetics and Genomics of Core Short Tandem Repeat Loci Used in Human Identity Testing. Journal of Forensic Sciences 2006;51:253–65. doi:10.1111/j.1556-4029.2006.00046.x.

[10] Jacewicz R, Jedrzejczyk M, Ludwikowska M, Berent J. Population database on 15 autosomal STR loci in 1000 unrelated individuals from the Lodz region of Poland. Forensic Science International: Genetics 2008;2:e1–3. doi:10.1016/j.fsigen.2007.08.002.

[11] Kim J, Edge MD, Algee-Hewitt BFB, Li JZ, Rosenberg NA. Statistical Detection of Relatives Typed with Disjoint Forensic and Biomedical Loci. Cell 2018;175:848-858.e6. doi:10.1016/j.cell.2018.09.008.

[12] Parsons TJ, Huel RML, Bajunović Z, Rizvić A. Large scale DNA identification: The ICMP experience. Forensic Science International: Genetics 2019;38:236–44. doi:10.1016/j.fsigen.2018.11.008.

[13] Egeland T, Sheehan N. On identification problems requiring linked autosomal markers. Forensic Science International: Genetics 2008;2:219–25. doi:10.1016/j.fsigen.2008.02.006.

[14] Pinto N, Gusmão L, Amorim A. X-chromosome markers in kinship testing: A generalisation of the IBD approach identifying situations where their contribution is crucial. Forensic Science International: Genetics 2011;5:27–32. doi:10.1016/j.fsigen.2010.01.011.

[15] Alzate LN, Agudelo N, Builes JJ. X-STRs as a tool for missing persons identification using only siblings as reference. Forensic Science International: Genetics Supplement Series 2015;5:e636–7. doi:10.1016/j.fsigss.2015.10.008.

[16] Kayser M. Forensic use of Y-chromosome DNA: a general overview. Human Genetics 2017;136:621–35. doi:10.1007/s00439-017-1776-9.

[17] Nakahori Y, Takenaka O, Nakagome Y. A human X-Y homologous region encodes "amelogenin." Genomics 1991;9:264–9. doi:10.1016/0888-7543(91)90251-9.
[18] Mannucci A, Sullivan KM, Ivanov PL, Gill P. Forensic application of a rapid and quantitative DNA sex test by amplification of the XY homologous gene amelogenin. International Journal of Legal Medicine 1994;106:190–193.
[19] Bär W, Brinkmann B, Budowle B, Carracedo A, Gill P, Holland M, et al. Guidelines for mitochondrial DNA typing. DNA Commission of the International Society for Forensic Genetics. Vox Sang 2000;79:121–5. doi:10.1159/000031227.
[20] Bourdon V, Ng C, Harris J, Prinz M, Shapiro E. Optimization of human mtDNA control region sequencing for forensic applications. J Forensic Sci 2014;59:1057–63. doi:10.1111/1556-4029.12426.
[21] Luo S, Valencia CA, Zhang J, Lee N-C, Slone J, Gui B, et al. Biparental Inheritance of Mitochondrial DNA in Humans. Proceedings of the National Academy of Sciences 2018;115:13039–44. doi:10.1073/pnas.1810946115.
[22] Watherston J, McNevin D, Gahan ME, Bruce D, Ward J. Current and emerging tools for the recovery of genetic information from post mortem samples: New directions for disaster victim identification. Forensic Sci Int Genet 2018;37:270–82. doi:10.1016/j.fsigen.2018.08.016.
[23] Sobrino B, Brión M, Carracedo A. SNPs in forensic genetics: a review on SNP typing methodologies. Forensic Science International 2005;154:181–94. doi:10.1016/j.forsciint.2004.10.020.
[24] Budowle B, van Daal A. Forensically relevant SNP classes. BioTechniques 2008;44:603–8, 610. doi:10.2144/000112806.
[25] Qiao L, Yang Y, Fu P, Hu S, Zhou H, Peng S, et al. Genome-wide variants of Eurasian facial shape differentiation and a prospective model of DNA based face prediction. Journal of Genetics and Genomics 2018;45:419–32. doi:10.1016/j.jgg.2018.07.009.
[26] Oldoni F, Kidd KK, Podini D. Microhaplotypes in forensic genetics. Forensic Sci Int Genet 2019;38:54–69. doi:10.1016/j.fsigen.2018.09.009.
[27] Fondevila M, Pereira R, Gusmão L, Phillips C, Lareu MV, Carracedo A, et al. Forensic performance of insertion–deletion marker systems. Forensic Science International: Genetics Supplement Series 2011;3:e443–4. doi:10.1016/j.fsigss.2011.09.083.
[28] Fondevila M, Phillips C, Naveran N, Fernandez L, Cerezo M, Salas A, et al. Case report: identification of skeletal remains using short-amplicon marker analysis of severely degraded DNA extracted from a decomposed and charred femur. Forensic Sci Int Genet 2008;2:212–8. doi:10.1016/j.fsigen.2008.02.005.
[29] Carnell AN, Goodman JI. The long (LINEs) and the short (SINEs) of it: altered methylation as a precursor to toxicity. Toxicol Sci 2003;75:229–35. doi:10.1093/toxsci/kfg138.
[30] Alvarez-Cubero MJ, Saiz M, Martinez-Gonzalez LJ, Alvarez JC, Eisenberg AJ, Budowle B, et al. Genetic Identification of Missing Persons: DNA Analysis of Human Remains and Compromised Samples. Pathobiology 2012;79:228–38. doi:10.1159/000334982.
[31] Ansell R. Internal quality control in forensic DNA analysis. Accreditation and Quality Assurance 2013;18:279–89. doi:10.1007/s00769-013-0968-9.
[32] Butler JM. The future of forensic DNA analysis. Philosophical Transactions of the Royal Society B: Biological Sciences 2015;370:20140252. doi:10.1098/rstb.2014.0252.
[33] M Romeika J. Recent Advances in Forensic DNA Analysis. Journal of Forensic Research 2014;s12. doi:10.4172/2157-7145.S12-001.
[34] Ward J. Best practice recommendations for the establishment of a national DNA identification program for missing persons: A global perspective. Forensic Science International: Genetics Supplement Series 2017;6:e43–5. doi:10.1016/j.fsigss.2017.09.009.
[35] Iwamura ESM, Guimarães MA, Evison MP. DNA Methods to Identify Missing Persons. In: Morewitz SJ, Sturdy Colls C, editors. Handbook of Missing Persons, Cham:

Springer International Publishing; 2016, p. 337–52. doi:10.1007/978-3-319-40199-7_22.
[36] Prinz M, Carracedo A, Mayr WR, Morling N, Parsons TJ, Sajantila A, et al. DNA Commission of the International Society for Forensic Genetics (ISFG): Recommendations regarding the role of forensic genetics for disaster victim identification (DVI). Forensic Science International: Genetics 2007;1:3–12. doi:10.1016/j.fsigen.2006.10.003.
[37] Bajželj M, Pajnič IZ. Missing persons genetic identification. Zdravniski Vestnik 2017;86.
[38] ISO. ISO 18385:2016 : Minimizing the risk of human DNA contamination in products used to collect, store and analyze biological material for forensic purposes -- Requirements 2016.
[39] Forensics Science Regulator. The Control and Avoidance of Contamination In Crime Scene Examination involving DNA Evidence Recovery FSR-G-206 2016.
[40] ENFSI DNA working group. DNA database management review and recommendations 2017.
[41] Jäger AC, Alvarez ML, Davis CP, Guzmán E, Han Y, Way L, et al. Developmental validation of the MiSeq FGx Forensic Genomics System for Targeted Next Generation Sequencing in Forensic DNA Casework and Database Laboratories. Forensic Science International: Genetics 2017;28:52–70. doi:10.1016/j.fsigen.2017.01.011.
[42] Parson W, Ballard D, Budowle B, Butler JM, Gettings KB, Gill P, et al. Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements. Forensic Sci Int Genet 2016;22:54–63. doi:10.1016/j.fsigen.2016.01.009.
[43] Steele CD, Balding DJ. Choice of population database for forensic DNA profile analysis. Science & Justice 2014;54:487–93. doi:10.1016/j.scijus.2014.10.004.
[44] Weir BS. Matching and partially-matching DNA profiles. J Forensic Sci 2004;49:1009–14.
[45] Ge J, Eisenberg A, Budowle B. Developing criteria and data to determine best options for expanding the core CODIS loci. Investig Genet 2012;3:1. doi:10.1186/2041-2223-3-1.
[46] Budowle B, Ge J, Chakraborty R, Gill-King H. Use of prior odds for missing persons identifications. Investigative Genetics 2011;2:15. doi:10.1186/2041-2223-2-15.
[47] Ge J, Budowle B, Chakraborty R. Choosing Relatives for DNA Identification of Missing Persons. Journal of Forensic Sciences 2011;56:S23–8. doi:10.1111/j.1556-4029.2010.01631.x.
[48] Vullo CM, Romero M, Catelli L, Šakić M, Saragoni VG, Jimenez Pleguezuelos MJ, et al. GHEP-ISFG collaborative simulated exercise for DVI/MPI: Lessons learned about large-scale profile database comparisons. Forensic Science International: Genetics 2016;21:45–53. doi:10.1016/j.fsigen.2015.11.004.
[49] Kling D, Tillmar AO, Egeland T. Familias 3 - Extensions and new functionality. Forensic Sci Int Genet 2014;13:121–7. doi:10.1016/j.fsigen.2014.07.004.
[50] van Dongen C, Slooten K, Slagter M, Burgers W, Wiegerinck W. Bonaparte: Application of new software for missing persons program. Forensic Science International: Genetics Supplement Series 2011;3:e119–e120.
[51] Butler JM. Forensic DNA advisory groups: dab, swgdam, enfsi, and bsag. Encyclopedia of Forensic Sciences, 2nd Edition 2012;1.
[52] Santos F, Machado H, Silva S. Forensic DNA databases in European countries: is size linked to performance? Life Sciences, Society and Policy 2013;9. doi:10.1186/2195-7819-9-12.
[53] Council of the European Union. Council Decision 2008/616/JHA of 23 June 2008 on the implementation of Decision 2008/615/JHA on the stepping up of cross-border cooperation, particularly in combating terrorism and cross-border crime. Official Journal of the European Union 2008.
[54] Huffine E, Crews J, Kennedy B, Bomberger K, Zinbo A. Mass identification of persons missing from the break-up of the former Yugoslavia: structure, function, and role of

the International Commission on Missing Persons. Croatian Medical Journal 2001;42:271–275.
[55] The International Commission on Missing Persons. ICMP Laboratory in The Hague Utilizes New Technology for Human Identification [Press Release] 2019.
[56] Lorente JA, Entrala C, Alvarez JC, Arce B, Heinrichs B, Lorente M, et al. Identification of missing persons: the Spanish "Phoenix" program. Croat Med J 2001;42:267–70.
[57] Gurkan C, Demirdov DK, Sevay H. Population genetics of Turkish Cypriots from Cyprus: Forensic and anthropological implications. Forensic Science International: Genetics Supplement Series 2015;5:e384–6. doi:10.1016/j.fsigss.2015.09.152.
[58] Committee on Enforced Disappearances (United Nations). Draft guiding principles for the search for disappeared persons 2018.
[59] Internationales Komitee vom Roten Kreuz, Interparlamentarische Union, editors. Missing persons. Geneva: IPU; 2009.
[60] AMBER Alert Europe. 5 key recommendations to save the lives of endangered missing children in Europe 2016.
[61] Hares DR. Selection and implementation of expanded CODIS core loci in the United States. Forensic Science International: Genetics 2015;17:33–4. doi:10.1016/j.fsigen.2015.03.006.
[62] Martin PD, Schmitter H, Schneider PM. A brief history of the formation of DNA databases in forensic science within Europe. Forensic Science International 2001;119:225–31. doi:10.1016/S0379-0738(00)00436-9.
[63] Council of the European Union. Council Resolution of 25 June 2001 on the exchange of DNA analysis results (2001/C 187/1) 2001.
[64] Council of the European Union. Council Resolution of 30 November 2009 on the exchange of DNA analysis results (2009/C 296/01) 2009.
[65] Alleyne L, Group IDME, Organization ICP. Interpol Handbook on DNA Data Exchange and Practice: Recommendations from the Interpol DNA Monitoring Expert Group. Interpol; 2009.
[66] International Committee of the Red Cross. Missing people, DNA analysis and identification of human remains. 2019.
[67] U.S. Department of Justice. Lessons Learned From 9/11: DNA Identification in Mass Fatality Incidents. 2006.
[68] Council of the European Union. Note 10900/05 from the Council Secretariat to the delegations regarding the Prum Convention 2005.
[69] Santos F. Overview of the implementation of the Prüm Decisions. Overview of the Implementation of the Prüm Decisions 2016:1–25.
[70] Council of the European Union. Council Decision 2008/615/JHA of 23 June 2008 on the stepping up of cross-border cooperation, particularly in combating terrorism and cross-border crime. Official Journal of the European Union 2008;210:1–11.
[71] Toom V. Cross-border exchange and comparison of forensic DNA data in the context of the Prüm decision. Brussels: European Parliament 2018.
[72] Council of the European Union. Note 7609/1/11 from the German delegation to the Working Group on Information Exchange and Data Protection on the Implementation of Council Decisions 2008/615/JHA and 2008/616/JHA n.d.
[73] Council of the European Union. Note 14918/10 from the Presidency to the Working Party on Information Exchange and Data Protection (DAPIX), on the Analysis of replies to the questionnaire on the Implementation of the "Prum Decisions" 2010.
[74] Council of the European Union. Note 5017/3/16 from the Presidency to the Working Party on Information Exchange and Data Protection (DAPIX) on the Implementation of the provisions on information exchange of the "Prum Decisions" 2016.
[75] Taverne MD, Broeders APA. Cross-border patterns in DNA matches between the Netherlands and Belgium. Sci Justice 2017;57:28–34. doi:10.1016/j.scijus.2016.08.008.
[76] Council of the European Union. Council conclusions on the vision for European Forensic Science 2020 including the creation of a European Forensic Science Area and the development of forensic science infrastructure in Europe 2011.

[77] Pádár Z, Nogel M, Kovács G. Accreditation of forensic laboratories as a part of the "European Forensic Science 2020" concept in countries of the Visegrad Group. Forensic Science International: Genetics Supplement Series 2015;5:e412–e413.
[78] ISO/IEC. ISO/IEC 17025: General requirements for the competence of testing and calibration laboratories 2017.
[79] Goor R, Riley G, Hoffman D. OSIRIS – Open Source Software STR Analysis Tool Expanded Process Control Capability (Poster) 2015.
[80] Bright J-A, Curran JM, Buckleton JS. The effect of the uncertainty in the number of contributors to mixed DNA profiles on profile interpretation. Forensic Science International: Genetics 2014;12:208–14. doi:10.1016/j.fsigen.2014.06.009.
[81] Roebuck H. Five quick facts about forensic mixed DNA profiles (LinkedIn post) 2015.
[82] Bieber FR, Buckleton JS, Budowle B, Butler JM, Coble MD. Evaluation of forensic DNA mixture evidence: protocol for evaluation, interpretation, and statistical calculations using the combined probability of inclusion. BMC Genetics 2016;17. doi:10.1186/s12863-016-0429-7.
[83] Budowle B, Onorato AJ, Callaghan TF, Manna AD, Gross AM, Guerrieri RA, et al. Mixture Interpretation: Defining the Relevant Features for Guidelines for the Assessment of Mixed DNA Profiles in Forensic Casework. Journal of Forensic Sciences 2009;54:810–21. doi:10.1111/j.1556-4029.2009.01046.x.
[84] Taylor D, Bright J-A, Buckleton J. The interpretation of single source and mixed DNA profiles. Forensic Sci Int Genet 2013;7:516–28. doi:10.1016/j.fsigen.2013.05.011.
[85] Van der Beek CP. Forensic DNA Profiles Crossing Borders in Europe (Implementation of the Treaty of Prüm) 2011.
[86] Lapointe M, Rogic A, Bourgoin S, Jolicoeur C, Séguin D. Leading-edge forensic DNA analyses and the necessity of including crime scene investigators, police officers and technicians in a DNA elimination database. Forensic Science International: Genetics 2015;19:50–55.
[87] ICMP. DNA Exclusion Database: Anonymous, Secure and Searchable (Press release) 2018.
[88] Gill P, Gusmão L, Haned H, Mayr WR, Morling N, Parson W, et al. DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods. Forensic Sci Int Genet 2012;6:679–88. doi:10.1016/j.fsigen.2012.06.002.
[89] Milot E, Courteau J, Crispino F, Mailly F. Inclusion probability with dropout: an operational formula. Forensic Sci Int Genet 2015;16:71–6. doi:10.1016/j.fsigen.2014.11.023.
[90] Mitchell AA, Tamariz J, O'Connell K, Ducasse N, Budimlija Z, Prinz M, et al. Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in. Forensic Sci Int Genet 2012;6:749–61. doi:10.1016/j.fsigen.2012.08.007.
[91] Amigo J, Phillips C, Salas T, Formoso LF, Carracedo Á, Lareu M. pop. STR—an online population frequency browser for established and new forensic STRs. Forensic Science International: Genetics Supplement Series 2009;2:361–362.
[92] Bodner M, Bastisch I, Butler JM, Fimmers R, Gill P, Gusmão L, et al. Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on quality control of autosomal Short Tandem Repeat allele frequency databasing (STRidER). Forensic Science International: Genetics 2016;24:97–102.
[93] Ruitberg CM, Reeder DJ, Butler JM. STRBase: a short tandem repeat DNA database for the human identity testing community. Nucleic Acids Research 2001;29:320–322.
[94] Cock PJA, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Research 2010;38:1767–71. doi:10.1093/nar/gkp1137.
[95] Parson W, Gusmão L, Hares DR, Irwin JA, Mayr WR, Morling N, et al. DNA Commission of the International Society for Forensic Genetics: Revised and

extended guidelines for mitochondrial DNA typing. Forensic Science International: Genetics 2014;13:134–42. doi:10.1016/j.fsigen.2014.07.010.
[96] ISO/IEC. ISO/IEC 19794-14:2013 Information technology -- Biometric data interchange formats -- Part 14: DNA data 2013.

List of abbreviations and definitions

CMP	Committee on Missing Persons
CODIS	Combined DNA Index System
CODIS CMF	CODIS Common Message Format
CRE	CODIS Rapid Enrolment
CS-SIS	Central System SIS
DAB	DNA Advisory Board
DNA	DeoxyriboNucleic Acid
DVI	Disaster Victim Identification
EAAF	Argentine Forensic Anthropology Team
EDNAP	European DNA Profiling
EFSA	European Food Safety Authority
ENFSI	European Network of Forensic Science Institutes
ESS	European Standard Set
FBI	Federal Bureau of Investigation
fDMS	Forensic Data Management System
HVI and HVII	Hypervariable Regions
IAC	Internal Amplification Control
ICMP	International Commission on Missing Persons
ICPO-INTERPOL	International Criminal Police Organization - INTERPOL
InDels	Insertion-deletion polymorphisms
ISFG	International Society for Forensic Genetics
ISSOL	Standard Set of Loci
LINEs	Long Interspersed Elements
LR	Likelihood Ratio
LSJML	Laboratoire de Sciences Judiciaires et de Médecine Légale
MED	Manufacturers' Exclusion Database
MPB	Missing Persons Bureau
MtDNA	Mitochondrial DNA
NIST	National Institute of Standards and Technology
OSAC	Organization of Scientific Area Committee
OSIRIS	Open Source Independent Review and Interpretation System
PCR	Polymerase Chain Reaction
programa FENIX	Spanish Phoenix program
QAS	Quality Assurance Standards
qPCR	Quantitative Polymerase Chain Reaction
SINEs	Short Interspersed Elements

SIS	Schengen Information System
SNP	Single-nucleotide polymorphism
sMIME	Secure/Multipurpose Internet Mail Extensions
STRs	Short Tandem Repeats
SWGDAM	Scientific Working Group of DNA Analysis Methods
TESTA	Trans European Services for Telematics between Administrations
TWGDAM	Technical Working Group on DNA Analysis Methods
VIS	Visa Information System
XML	eXtensible Markup Language

List of figures

Figure 1. Present functionality of CS-SIS in the scope of the recently adopted regulation [3]..... 7

Figure 2. Visual illustration of the biology and inheritance of STRs. (A) Marker D18S51 consists of repeats of the sequence AGAA (yellow rectangle). In this example, one allele contains 10 of these repeats, the second allele 14. (B) For each chromosome, parents contribute one of their two copies to their offspring, selected randomly in the genesis of the gametes. This schema illustrates one of the possible combinations (for two markers), given by the parents to their children.....11

Figure 3. Relatives from which Y-STR information could be informative to identify a missing person (circled), if - and only if - the missing person is male. 13

Figure 4. Relatives from which mtDNA information could be informative to identify a missing person (circled), if the missing person is male (A) or female (B). 14

Figure 5. Example family tree as part of a DNA reference profile collection form. The donor is invited to circle their relationship to the missing individual. 35

Figure 6. Example of a DNA profile as generated for forensics purposes, including (A) the repeat values of different STR markers and (B) mtDNA sequence of the HV1 region 40

Figure 7. Example electropherogram of an experiment to generate a DNA profile, which is interpreted from comparing the size of the alleles at the different markers with standard size ladders and allelic cocktail.....41

Figure 8. Schematic representation of the structure and fields of the Prüm XML format. Highlighted are (1) the field for the type of profile ("person" or "stain") and (2) the fields to include the information about the specified set of DNA markers. 47

Figure 9. Schematic representation of the structure and fields of the CODIS XML formats. A: the CODIS CMF version 4.3 ; highlighted are (1) the description of the source of the profile (specimen) and (2) the information about the markers analysed for each specimen. B: the CODIS Rapid Import CMF Version 1.0; highlighted are (1) the information on the device used to generate the profile, (2) the description of the source of the profile (specimen) and (3) the information about the markers analysed for each specimen.....48

Figure 10. Schematic representation of the structure and fields of the ISO 19794-14 XML format. Highlighted are (1) the full pedigree related to the profile(s) included, (2) the certification of the laboratory that generated the profile, (3) the information about the markers analysed for each sample, if they are STRs and (4) sequence information, if the marker used is mtDNA.....49

List of tables

Table 1. Maximum amounts of typically recoverable DNA from the various biological materials used in forensic sciences 19

Table 2: Summary of the STR loci present in the different published standard sets, including the additional loci included in INTERPOL's DNA profile information form 34

Table 3. Range of reported values for each allele in databases compiling information about the frequencies in different populations. The last column shows the upper and lower limits for each marker that can be used to compare the values in a final DNA profile 44

Table 4. Range of reported values for commonly used Y-STR alleles from the STRidER database 45

Annex

Annex 1. List of relevant guidelines and recommendations documents produced by the working groups and organisations described in the report.

ENFSI

- Recommended Minimum Criteria for the Validation of Various Aspects of the DNA Profiling Process (2010) - recommendations on how to validate methods used in laboratories generating DNA profiles
- ENFSI Survey on the DNA Profile Inclusion, Removal and Retention of Member Sates' Forensic DNA Databases: Chris Asplen (2009) - a survey of the criteria for DNA profile inclusion, retention and exclusion from different national DNA databases
- DNA database management review and recommendations (2017) - recommendations on different aspects of forensic DNA database management

DAB

- Statistical and Population Genetics Issues Affecting the Evaluation of the Frequency of Occurrence of DNA Profiles Calculated from Pertinent Population Database (2000) - a statistical guide to identify or exclude an individual(s) as a contributor(s) to a specific DNA profile
- Quality Assurance Standards Audit for DNA Databasing Laboratories (2011) - an audit document for assessing compliance with the DAB forensic DNA testing and DNA databasing laboratories standards

ISFG

- DNA Commission of the International Society for Forensic Genetics (ISFG): Recommendations regarding the role of forensic genetics for disaster victim identification (DVI) (2007) - recommendations for forensic genetics laboratories on collecting and storing DNA profiles in the aftermath of large-scale disasters
- Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on quality control of autosomal Short Tandem Repeat allele frequency databasing (STRidER) (2016) - guidelines for the inclusion of data into the STRidER database
- ISFG: Recommendations on biostatistics in paternity testing (2007) – guidelines for calculating strength of evidence in kinship cases.
- DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis (2006) - Recommendations regarding the nomenclature, the definition of loci and alleles, population genetics and reporting methods for Y-STRs
- DNA Commission of the International Society for Forensic Genetics (ISFG): Guidelines on the use of X-STRs in kinship analysis - Same, for X-STRs
- DNA Commission of the International Society for Forensic Genetics: revised and extended guidelines for mitochondrial DNA typing – same, for mitochondrial DNA.

SWGDAM

- SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories (2017)- guidelines for the interpretation of DNA typing results from short tandem repeats (STR)
- SWGDAM Guidelines for Missing Persons Casework (2014) - guidelines for the processing and analysis (DNA recovery, reference samples, etc.) of missing persons casework

- SWGDAM Validation Guidelines for DNA Analysis Methods (2016) - guide for laboratories in validating procedures consistent with the FBI Director's Quality Assurance Standards (QAS)

ICPO-Interpol

- Best Practice Principles: Recommendations for the Establishment of a National DNA Database (2015) - recommendations for INTERPOL member countries wishing to establish a national DNA database
- Best Practice Principles: Recommendations on the Use of DNA for the Identification of Missing Persons and Unidentified Human Remains (2015) - recommendations for INTERPOL member countries wishing to use DNA for the identification of missing persons and unidentified human remains in police investigations

Note: All the links were last accessed on 10 April 2019.

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: http://europea.eu/contact

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

- at the following standard number: +32 22999696, or

- by electronic mail via: http://europa.eu/contact

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: http://europa.eu

EU publications
You can download or order free and priced EU publications from EU Bookshop at: http://bookshop.europa.eu. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see http://europa.eu/contact).

KJ-NA-29766-EN-N

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub

Publications Office

doi:10.2760/13343

ISBN 978-92-76-07983-5