A peer-reviewed version of this preprint was published in PeerJ on 13 August 2019.

View the peer-reviewed version (peerj.com/articles/7314), which is the preferred citable publication unless you specifically need to cite this preprint.

Mitochondrial DNA in human identification: a review

António Amorim 1, Teresa Fernandes 2, Nuno Taveira 3, 4

1 Instituto Nacional de Medicina Legal e Ciências Forenses, Lisbon, Portugal
2 Biologia, Universidade de Évora, Évora, Portugal
3 Instituto Universitário Egas Moniz (IUEM), Almada, Portugal
4 Research Institute for Medicines (iMed.ULisboa), Faculty of Pharmacy, Universidade de Lisboa, Lisbon, Portugal

Corresponding Author: Nuno Taveira
Email address: ntaveira@ff.ul.pt

Mitochondrial DNA (mtDNA) presents several characteristics useful for forensic studies, especially related to the lack of recombination, to a high copy number, and to matrilineal inheritance. mtDNA typing based on sequences of the control region or full genomic sequences analysis is used to analyze a variety of forensic samples such as old bones, teeth and hair, as well as other biological samples where the DNA content is low. Evaluation and reporting of the results requires careful consideration of biological issues as well as other issues such as nomenclature and reference population databases. In this work we review mitochondrial DNA profiling methods used for human identification and present their use in the main cases of human identification focusing on the most relevant issues for the forensic and medico-legal areas.
Mitochondrial DNA in Human Identification: a review

Authors

António Amorim¹, Teresa Fernandes², Nuno Taveira³,⁴*

Affiliations

¹Instituto Nacional de Medicina Legal e Ciências Forenses, Lisboa, Portugal
²Universidade de Évora, Évora, Portugal
³Instituto Universitário Egas Moniz (IUEM), Almada, Portugal
⁴Research Institute for Medicines (iMed.ULisboa), Faculty of Pharmacy, Universidade de Lisboa, Portugal

* Corresponding author:

Nuno Taveira

Research Institute for Medicines (iMed.ULisboa), Faculty of Pharmacy, Universidade de Lisboa, Av. Prof. Gama Pinto, 1649-003 Lisbon, Portugal

ntaveira@ff.ul.pt
Abstract

Mitochondrial DNA (mtDNA) presents several characteristics useful for forensic studies, especially related to the lack of recombination, to a high copy number, and to matrilineal inheritance. mtDNA typing based on sequences of the control region or full genomic sequences analysis is used to analyze a variety of forensic samples such as old bones, teeth and hair, as well as other biological samples where the DNA content is low. Evaluation and reporting of the results requires careful consideration of biological issues as well as other issues such as nomenclature and reference population databases. In this work we review mitochondrial DNA profiling methods used for human identification and present their use in the main cases of human identification focusing on the most relevant issues for the forensic and medico-legal areas.
Introduction

Human genetic identification for medico-legal or forensic purposes is achieved through the definition of genetic profiles. A genetic profile or the genetic fingerprint of an individual is the phenotypic description of a set of genomic loci that are specific to that individual. In accordance with international recommendations, particularly with recommendations of the European DNA Profiling Group (EDNAP), currently, only genetic profiles obtained from autosomal short tandem repeats (STR) should be used for genetic fingerprinting. However, in a considerable number of situations of medico-legal and forensic identification, autosomal DNA is highly degraded or isn’t available at all. In these cases the study of mitochondrial DNA (mtDNA) for human identification has become routine (1–5).

Mitochondrial DNA (mtDNA) presents several characteristics useful for forensic studies, especially related to the lack of recombination, to a high copy number, and to matrilineal inheritance. mtDNA typing based on sequences of the control region or full genomic sequences analysis is used to analyze a variety of forensic samples such as old bones, teeth and hair, as well as other biological samples where the DNA content is low. Evaluation and reporting of the results requires careful consideration of biological issues as well as other issues such as nomenclature and reference population databases. In this work we review mitochondrial DNA profiling methods used for human identification and present their use in the main cases of human identification focusing on the most relevant issues for the forensic and medico-legal areas.
We systematically searched in PubMed for papers in English describing 1) mitochondrial DNA biology and genetics, 2) mitochondrial DNA typing guidelines, 3) mitochondrial DNA nomenclature, 4) mitochondrial DNA sequencing methodologies, 5) mitochondrial DNA population data and databases and 6) mitochondrial DNA in human identification and forensics. Our search wasn’t refined by publishing date, journal or impact factor of the journal, authors or authors affiliations. In addition, we used Guideline documents from the International Society for Forensic Genetics available at [https://www.isfg.org/](https://www.isfg.org/).

Mitochondrial DNA biology and genetics

Mitochondria are cellular organelles that contain an extrachromosomal genome, which is both different and separate from the nuclear genome. The mitochondrial DNA (mtDNA) was first identified and isolated by Margit Nass and Sylvan Nass in 1963, who studied some mitochondrial fibers that according to their fixation, stabilization and staining behavior, appeared to be DNA related (6). However, the complete sequence of the first mtDNA was only published and established as the mtDNA Cambridge Reference Sequence (CRS) eighteen years later, in 1981 (7).

Essentially, the mtDNA is a 5 mm histone-free circular double-stranded DNA molecule, with around 16,569 base-pairs and weighting $10^7$ Daltons (8). mtDNA strands have different densities due to different G+T base composition. The heavy (H) strand encodes more information, with genes for two rRNAs (12S and 16S), twelve polypeptides and fourteen tRNAs, while the light (L) strand encodes eight tRNAs and one polypeptide. All the 13 protein products are part of the enzyme complexes that constitute the oxidative phosphorylation system. Other characteristic
features of the mtDNA are the intronless genes and the limited, or even absent, intergenic sequences, except in one regulatory region.

The mitochondrial D-loop is a triple-stranded region found in the major non-coding region (NCR) of many mitochondrial genomes, and is formed by stable incorporation of a third 680 bases DNA strand known as 7S DNA (9). The origin of replication is located at the non-coding or D-loop region, a 1121 base pairs segment that is located between positions 16,024 and 516, according to the CRS numeration (7) (Figure 1). The D-loop region, also comprehends two transcription promotors, one for each strand. Nucleotide positions in the mtDNA genome are numbered according to the convention presented by Anderson et al. (7), which was slightly modified by Andrews et al. (10). More precisely, the numerical designation of each base pair is initiated at an arbitrary position on the H strand, which continues thereafter and around the molecule for approximately 16,569 base pairs.

The apparent lack of mtDNA repair mechanisms and the low fidelity of the mtDNA polymerase lead to a significant higher mutation rate in the mitochondrial genome, when compared to the nuclear genome. For example, Sigurðardóttir and collaborators, estimated the mutation rate in the human mtDNA control region to be $0.32 \times 10^{-6}$/site/year (10) which compares to $0.5 \times 10^{-9}$/site/year in the nuclear genome (12). Most of the sequence variation between individuals is found in two specific segments of the control region, namely in the hypervariable region 1 (HV1, positions 16,024 to 16,365) and in the hypervariable region 2 (HV2, positions 73 to 340) (13). A third hypervariable region (HV3, positions 438 to 574), with additional polymorphic positions can be useful in the resolution of indistinguishable HV1/HV2 samples (14). The small size and relatively high inter-person variability of the HV regions are very useful features for forensic testing purposes.
A mitochondrion contains 2 to 10 copies of mtDNA and each somatic cell can have up to 1,000 mitochondria (15,16). Hence, when the amount of the extracted DNA is quite small or degraded, it is more likely that a DNA typing result can be obtained by typing the mtDNA than by typing polymorphic regions that are found in nuclear DNA.

Contrarily to the nuclear DNA, the mtDNA is exclusively maternally inherited, which justifies the fact that, apart from mutation, mtDNA sequence of siblings and all maternal relatives is identical (17–19). This specific characteristic can be very helpful in forensic cases, such as in the analysis of the remains of a missing person, where the known maternal relatives can provide some reference samples for a direct comparison to the mtDNA type. Due to the lack of recombination, maternal relatives from several generations apart from the source of evidence (or biological material) can be used for reference samples (17–19).

The haploid and monoclonal nature of the mtDNA in most individuals simplifies the process of interpretation of the DNA sequencing results. Still, it is possible to find heteroplasmy at occasional cases (20–25). A person is considered as heteroplasmic if she/he carries more than one detectable mtDNA type. There are two classes of heteroplasmy, related to length polymorphisms and to point substitutions. Only the latter is important for forensic human identification. Most forensic laboratories worldwide do not report length polymorphisms and the guidelines on human identification with mtDNA do not point them as mandatory information (2,4). Furthermore, the information of length polymorphisms has no impact in haplogroups’ definition.

Heteroplasmy manifests itself in diverse ways (26). An individual may show more than one mtDNA type in a single tissue. An individual may be heteroplasmic in one tissue sample and homoplastic in another one. Finally, an individual may exhibit one mtDNA type in one tissue
and a different type in another tissue. Of the three possible scenarios, the last one is the least likely to occur. When heteroplasmy is found in the mtDNA of an individual, it usually differs at a single base, in HV1 or HV2.

Heteroplasmy was observed at position 16,169 of the mtDNA control region in the putative remains of Tsar Nicholas II of Russia and his brother, the Grand Duke of Russia Georgij Romanov (22,23). Comas et al. (21), in turn, detected heteroplasmy at two distinct positions, 16,293 and 16,311, in the mtDNA of an anonymous donor’s plucked hair. Wilson et al. (24) found a family constituted by a mother and two children carrying a heteroplasmic mtDNA at position 16,355 both in blood and buccal swab samples.

The existence of heteroplasmic individuals and the limited knowledge about both the mechanism and the rate of heteroplasmy can be issues raised in an attempt to exclude mtDNA evidence from forensic investigations. In the context of forensic analysis, both mtDNA sequences of a reference sample and an evidence sample(s) are compared. If the mtDNA sequences are identical, the samples can’t be excluded since they must have the same origin or derive from the same maternal lineage. Similarly, samples can’t be excluded when heteroplasmy is observed at the same nucleotide positions in both samples. Finally, when one sample is heteroplasmic and the other is homoplasmic but they both share at least one mtDNA species, the samples can’t be excluded since they may have the same origin. Several authors have suggested that samples with mtDNA with one-base difference should be further evaluated, mainly regarding their rate of mutation (1,27–29). When two or more nucleotide differences exist between the two sequences, the overall interpretation is exclusion (4).

Heteroplasmy at one nucleotide position is more frequently observed in hair samples, mainly due to genetic drift and to bottlenecks which occur due to the hair follicle’s semiclonal nature (5,30–
32). Hence, if an evidentiary hair sample contains one of the two heteroplasmic lineages that are observed in a reference sample, or vice versa, then the interpretation of exclusion may be incorrect. In this case, typing additional hairs may be required to solve the problem (5).

As it was previously pointed out, the mitochondrial genome is maternally inherited. Even though the sperm contains a few mitochondria in the neck and in the tail region, the male mitochondrial genome is destroyed either during or shortly after the fertilization. More precisely, sperm mitochondria disappear in the early embryogenesis, namely by selective destruction, inactivation or dilution (33–35). Nevertheless, there are a few examples of paternal inheritance of the mitochondrial genome in animals (36), and by that reason and despite the limited evidence for paternal inheritance of the mitochondrial genome in humans, the courtroom can use such possibility to exclude the use of mtDNA evidence.

Mitochondrial DNA Nomenclature

Even though the process of naming mtDNA sequences seems simple and obvious, it is crucial to properly consider the nomenclatures, since complications might arise. Considering that listing more than 600 bases in order to describe the results from a new HV1 and HV2 sequence would be unpractical, an alternative approach was developed which essentially identifies and reports the differences relative to a reference sequence (Cambridge Reference Sequence or CRS sequence) (7). For example, if the bases beyond the position 16,192 were out of the register by one base due to the insertion of a C the mutation is designated as 16,192.1 C. If two Cs were inserted, they would be designated as 16,192.1 C and 16,192.2 C. Regarding deletions, these are recorded by the number of the base(s) that is missing, with respect to the CRS (i.e., 249 D or 249-). The bases that cannot be unambiguously determined are coded by using an N. The guidelines that are used
to record all the sequence’s differences have been described by several authors (1,2,4,29), and
are used by the forensic community in general.

In forensics, the origin of the evidentiary sample is unknown. Consequently, and to be fair to the
defendant, it is important to list any sequences in a reference mtDNA database that are identical
to the evidentiary mtDNA sequence, especially when estimating how rare or common the
evidentiary mtDNA sequence is. Currently, most ambiguities in the alignment/nomenclature
arise due to insertions and/or deletions (indels). Hence, Wilson et al. (37) developed an approach
to attempt to standardize all mtDNA alignments which is based on a phylogenetic context, and
gives differential weighting to transversions, transitions, insertions and deletions.

The nomenclature approach mentioned above aims the standardization of reference mtDNA
databases, which is why recommendations are provided to address several scenarios (1,2,4,29).
The recommendations are as follows: a) characterize the variant(s), namely by using the least
number of differences (i.e., insertions, substitutions, deletions) from the reference sequence; b) if
there is more than one alignment, each one having the same number of differences when it
comes to the reference sequence, indels must be prioritized; c) indels must be placed at 3’ end,
especially with respect to the light strand (if possible, deletions and/or insertions must be
combined); and d) gaps are combined only when they can be placed at the 3’ end, while
maintaining the exact same number of differences from the reference sequence. The bases that
are designated with an N should not affect these recommendations. These recommendations are
hierarchical, which is why the first recommendation takes precedence, being immediately
followed by the remaining ones, that are, and to some degree, arbitrary. The last recommendation
is a clarification, since it aims to facilitate the alignment of all interpretations. Lastly, it is
relevant to point out that several examples have been identified, where alternative alignment
strategies are possible, being adequately described elsewhere in the literature (37).

Mitochondrial DNA Typing Guidelines

In 2014 the DNA Commission of the International Society of Forensic Genetics (ISFG)
published updated guidelines and recommendations concerning mitochondrial DNA typing.
These guidelines referred to good laboratory practices, targeted region, amplification and
sequencing ranges, reference sequence, alignment and notation, heteroplasmy, haplogrouping of
mtDNA sequences, and databases and database searches. In Table 2 we present the 16
recommendations of ISFG. Overall, these are the main guidelines concerning the application of
mtDNA polymorphisms in human identification, which are regularly revised and published by
the International Society of Forensic Genetics (1,2,4,29).

Mitochondrial DNA Sequencing Methodologies

In 1977 Sanger presented the first DNA sequencing technology (38), also called the chain
termination method and now known as first generation sequencing. The incorporation of ddNTPs
in newly synthesized DNA strands results in termination of the elongation process and
correspondent knowledge about the specific nucleotide present at the sequence at each position.
Sanger sequencing method can produce reads from 25 up to 1200 nucleotide positions, allowing
the read of a maximum of 96 kb nucleotides in 2 hours.
Since 2005 a wide number and variety of new sequencing methods have been developed and
launched on the market (39). Second generation sequencing methods do not allow to read the
complete sequence of an individual genome at one-time step but only small DNA fragments each
time, from 35 up to 75 nucleotide positions using the SOLiD technology (39–41), to 100 up to 1,000 nucleotide positions using the 454 Pyrosequencing technology (39,42). Depending on the technology, these methods allow sequence reads from 60 to 80 million nucleotide positions in 2 hours or 6 billion nucleotide positions in 1-2 weeks as (43,44).

Third generation sequencing technologies were introduced in the market in 2010 and allow massive parallel analysis combined with reading in real time (39). With these methods it is possible to read between 8,000 to 200,000 nucleotide positions at one-time, and they allow sequence reads from 100,000 up to 4 Tb nucleotide positions in less than 48 hours (45,46). Many different methodologies designed to attend to different priorities had been proposed. Some methodologies are focused on the number of reads, other’s give priority to results accuracy and other’s attend to the range of the reads. Pacific Biosciences (PacBio) introduced equipment’s designed to produce low number of reads with high accuracy (45,47), such as PacBio RSII (48), while Oxford Nanopore Technologies (ONT) designed methodologies to produce high number of reads with lower accuracy, such as ONT MinION (46), and Illumina gives priority to short-read with high accuracy technologies and present’s a ‘post-light’ methodology with directly sensing non-optical sequencing, such as Illumina HiSeq (49,50). These massive parallel sequencing (MPS) technologies have been quickly applied in a wide range of areas where forensics is included (39). Personal Genome Machine (PGM) first introduced the Ion Torrent methodology for sequence complete mtGenomes in forensic context (51). Nevertheless, regarding to mtDNA analysis for forensic human identification, and according to current international guidelines (2,4), Sanger sequencing still continues to be an adequate method and is used in most casework laboratories worldwide (52). However, if guidelines on human identification with mtDNA come to establish as mandatory the inclusion of information of the
coding region, the benefits of MPS could be explored to type whole mtDNA genomes. In 2015, Magalhães and collaborators presented the results of Ion Torrent PGM NGS technology applied to whole mtDNA molecule sequencing. Although it proved to be sensitive and accurate at detecting and quantifying mixture and heteroplasmy, there were some problems in the coverage of the mtDNA genome with some regions presenting extreme strand bias, and presenting false positives mostly generated by alignment problems in the analysis algorithms (53). Woerner and collaborators presented, in 2018, the evaluation of the Precision ID mtDNA Whole Genome Panel on two massively parallel sequencing commercial systems - Ion S5 System (Thermo Fisher Scientific) and MiSeq FGx Desktop Sequencer (Illumina) (54). According to their conclusions, Ion and MiSeq methodologies provide consistent mtDNA haplotypes estimation. Beyond this study many other studies on the use of MPS technologies for forensic genetics and mtDNA analysis have been published (55,56,65–67,57–64). However, further validation studies and specialized software functionality tailored to forensic practice should be produced in order to facilitate the incorporation of NGS processing into standard casework applications (68,69).

**Mitochondrial DNA Population Data and Databases**

When two mtDNA sequences, one from an evidence sample and another from a reference sample, cannot be excluded as being originated from the exact same source, it is necessary to convey some information concerning the rarity of the mtDNA profile. The current practice is to count how many times a specific sequence is observed within a population database(s) (70). Overall, the population databases that are used in forensics comprehend several convenience samples, representing the major population groups of the potential contributors in terms of evidence.
Phylogenetic methods have been used in order to identify the main human haplogroups, as well as the most important SNPs that define such groups. Regarding the construction of a phylogenetic tree of mtDNA sequences, there are several alternative methods, which include neighbor-joining, maximum likelihood, minimum spanning networks, maximum parsimony and Bayesian methods (71). Overall, the large majority of individuals from African populations, and specially from sub-Saharan African populations, are categorized into one of the main haplogroup lineages that diverged from macro-haplogroup L - L0, L1, L2, L3, L4, L5 and L6 - (72–79). On the other hand, more than 90% of the individuals of the European and USA Caucasian populations are categorized into 10 main haplogroup lineages - H, I, J, K, M, T, U, V, W and X - (5,76,77,80,81). Concerning to African-American populations, the most commonly observed haplogroups are L2a, L1c, L1b and L3b (75). The main haplogroups found in individuals from Asian populations are haplogroups M and N (82,83).

The first database for human mtDNA was Mitomap in 1995 (84). In 1996, this database developed into an online database, www.mitomap.org, containing published human mtDNA variation along with geographic and disease specific variants. Currently, Mitomap is manually curated, frequently updated and a functionally rich resource, presenting high-quality human mtDNA data for clinicians, investigators and geneticists (84). Mitomap has three main categories for usage. It contains some background information regarding the human mitochondrial DNA, such as the general representation of mtDNA, haplogroups and their frequencies and illustrations of mtDNA, among others. Furthermore, users can also find information about other mtDNA-specific databases, tools and useful resources. Mitomap stores the annotated listing of the mtDNA variants from both healthy individuals and patients. The frequencies of the variants are calculated from human mitogenomes retrieved from
the GenBank. Therefore, users can retrieve information about the loci, the nucleotide change, the
codon position and the number, among others, and download the most important data in different
file formats.

Mitomap contains the Mitomaster analysis tool, currently providing the Application
Programming Interface for it. The main function of this tool is to allow the identification of
polymorphic positions, the calculation of variant statistics and the assignment of haplogroups to
complete or partial mitogenomes. Such query might be performed by recurring to mtDNA
sequences, to GenBank identifiers or to single nucleotide variants (85).

Another database for human mtDNA is the EDNAP Mitochondrial DNA Population Database
(EMPOP, www.empop.org) (86). In its early stages, EMPOP was designed and envisioned to
serve as a reference population database, specifically to be used in the evaluation of the mtDNA
evidence around the world, aiming to provide the highest quality mtDNA data. The architecture
of this online database and its analysis tools, which are also provided via the website, have
evolved over the last few years, even though the main emphasis of the EMPOP database remains
to be mtDNA data quality. Therefore, and as a direct consequence, EMPOP not only serves as a
reference population database, but also as a quality-control tool for scientists in forensic genetics,
as well as in other disciplines. Finally, and even though there is a significant number of high-
quality reference population databases for forensic comparisons, EMPOP is the most
comprehensive resource, especially from the standpoint of the populations that are represented in
such database (4).

EMPOP uses SAM, a string-based search algorithm that converts query and database sequences
into alignment-free nucleotide strings and thus guarantees that a haplotype is found in a database
query regardless of its alignment. SAM-E, an updated version of SAM that considers block
InDels as phylogenetic events, is used currently. At EMPOP, the tool haplogroup browser represents all the established PhyloTree haplogroups in convenient searchable format and provides the number of EMPOP sequences assigned to the respective haplogroups by estimating mitochondrial DNA haplogroups using a maximum likelihood approach EMMA (87). For multiple possible haplogroups, most recent common ancestor (MRCA) haplogroups are provided.

In order to facilitate a better use of known mtDNA variation, van Oven & Kayser, in 2008, have constructed an updated comprehensive phylogeny of global human mtDNA variation - PhyloTree -, based on both coding and control region mutations (78). The complete mtDNA tree includes previously published as well as newly identified haplogroups, is continuously and regularly updated, and is available online at http://www.phylotree.org. In figure 2 we present the basic structure of the PhyloTree, which is divided into 25 subtrees accessible through http://www.phylotree.org. At EMPOP the geographical haplogroup patterns are provided via maps to visualize and better understand their geographical distribution (Figure 3).
From another perspective, ethical and legal problems may arise in the implementation of mtDNA databases. The informative potential which the analysis of mtDNA entails can generate privacy questions (88,89). Mitochondrial diseases affect between 1 in 4,000 and 1 in 5,000 people. In most people, primary mitochondrial disease is a genetic condition that can be inherited. Information about the mitochondrial genome composition may therefore enable the identification of the current or future state of health of an individual. For this reason, the analysis of mtDNA must be carried out only on non-coding regions, which have not been associated with any kind of disease or phenotypical information.

Mitochondrial DNA in Medico-Legal Human Identification

At this section we present some selected published cases of human identification with mtDNA. Table 1 summarizes the selected published cases.

In 1991, Stoneking and collaborators presented the first report of successful application of the mtDNA typing to a case that involved the individual identification of skeletal remains (90). This was the case of a 3-year-old child disappeared from her parents’ house in October of 1984. In March of 1986, the skeletal remains of a human child were found in the desert, 2 miles away from the parents’ residence. Using hybridization with 23 sequence-specific oligonucleotide probes (SSO) targeting nine regions of HV1 and HV2 on the control region, they found that the skeletal sample and the mother shared the same mtDNA types, corroborating that those skeletal remains were of the missing child. Moreover, they anticipated that the mtDNA typing would be valuable not only in linking biological remains to missing individuals, but also in the analysis of material in sexual assault cases.
In July of 1990, the body of a female, in a quite advanced state of decomposition, was discovered in an open field. Despite being impossible to identify the remains by analyzing the individual’s clothes and fingerprints, her dentition was consistent with old dental records of a missing person from the same region. Some fragments of the heel bone and fibula, plus samples of the hair and skin, were provided for the DNA analysis, as well as a blood sample from a putative sister of the deceased. In 1992, Sullivan and collaborators attempted the identification of the highly decomposed remains of the corpse, amplifying and directly sequencing 2 hypervariable segments within HV1 and HV2 in the mtDNA (91). No statistical value was given to the evidence, since no database of the British population sequences were available at that time. Still, no differences were found between both sequences, the blood of the putative sister and the bone of the corpse, indicating they were sisters.

Perhaps the most well-known lineage study using mtDNA sequencing is related to the identification of Tsar Nicholas II’s bones. Gill and collaborators, in 1994 (22), and Ivanov and collaborators, in 1996 (23), compared the sequences of HV1 and HV2 fragments of the mtDNA obtained from the putative bones of the Tsar with those of Tsar living maternal relatives, Countess Xenia Cheremeteff-Sfiri and the Duke of Fife. It was found that the sequences were very similar, corroborating the hypothesis that the bone remains were of Tsar Nicholas II.

In a distinct scenario, Deng et al. (92) used direct sequencing of the HV1 and HV2 fragments of the mtDNA control region to identify Tsunami victims in Thailand in 2004. This tsunami killed nearly 5,400 people in Southern Thailand, including foreign tourists and local residents. They succeeded in obtaining fully informative results for mtDNA markers (HV1 and HV2) from 258 tooth samples with a success rate of 51% (258/507).
More recently, in 2010, Ríos and collaborators (93) used direct sequencing of the HV1 and HV2 fragments of the mtDNA control region to identify human skeletal remains that were exhumed from a mass grave from the Spanish Civil War (1936-1939). There was a match between the mtDNA profiles of the biologically youngest skeleton and the sister of the youngest person that was presumptively known to be buried in the grave, allowing the identification of that person.

Also in 2010, Piccinini and collaborators (94) attempted to identify the remains of a famous World War One Italian soldier that was killed in a battle along the Italian front in 1915. Like previous studies, they used the direct sequencing of the HV1 and HV2 fragments of the mtDNA control region to define single mtDNA haplotypes. The availability of the offspring maternal lineage allowed the mtDNA analysis, which presented a clear exclusion scenario: the remains did not belong to the supposed war hero.

In 2012, a skeleton was excavated at the site of the Grey Friars friary, in Leicester, which is the last-known resting place of King Richard III (95). To determine if the remains belonged to King Richard III, the HV1, HV2 and HV3 regions of the mtDNA of the skeletal remains and of the living relatives of King Richard III were sequenced and compared. There was a perfect match between the sequences indicating that the remains belong to King Richard III.

The communist period in Poland during 1944-1956 resulted in the death of more than 50,000 people, who were buried in secret. One mass grave was found at the cemetery Powazki Military, in Warsaw, Poland. In 2016, Ossowski and collaborators (96) identified 50 victims, specifically by using autosomal, Y-STR and direct sequencing of the HV1 and HV2 fragments of the mtDNA control region.

In 2016, among the first studies on human identification with mtDNA using massive parallel sequencing, Park and collaborators proposed a protocol that includes the study of ten regions of
mtDNA for the identification of historical human remains with forensic genetic markers (97). They studied a 140-year-old human skeletal remains discovered at a historical site in Deadwood, South Dakota, United States. The remains were in an unmarked grave and there were no records available regarding the identity of the individual. The mtDNA profiles of the unidentified skeletal remains obtained with their method were consistent with H1 haplogroup. This haplogroup is the most common in Western Europe. The ancestry-informative nuclear SNPs also studied in this case indicated a European background. These genetic results are consistent with the findings of previous anthropological report which determined that the Deadwood unidentified skeletal remains belong to a male of European ancestry.

In 2017, the victims’ remains from the World Trade Center terrorism act, which occurred in September 11 of 2001, were still being identified by using the mtDNA sequencing technology, among other techniques, with protocols and guidelines as recommended by the International Society for Forensic Genetics (1,2,4,29,98).

Conclusions

Over the last 25 years, mtDNA typing has been widely used around the world to solve several human identifications related issues in violent crimes, lesser crimes, acts of terrorism, mass disasters and missing persons’ cases.

Forensic DNA methods are constantly questioned in terms of their admissibility for several years now, and we foresee that this scenario is likely to continue in the future. Some of the most well-known challenges to the mtDNA analysis are focused on the issue of the heteroplasmy as well as on recombination and paternal leakage.
Over the last decades progress in mtDNA typing was overwhelming, going from the examination of small fragment in a matter of days to sequencing multiple mtDNA genomes in a couple of hours using a point of care sequencing machine. An individual’s mtDNA genome can tell us much about his/her ancestors. Even though many would readily accept that there are good reasons for researchers to determine information about an unknown suspect’s potential ancestral background, many still might find the potential to determine genetic dispositions to certain disorders as being unacceptable. Hence, new technologies must be wisely used and for the reasons that they are intended, considering their specific focus and contribution within the field of forensic and medico-legal human identification.

Acknowledgements
None to report.

References


Ondov BD, Cochran C, Landers M, Meredith GD, Dudas G, Bergman NH. An alignment
algorithm for bisulfite sequencing using the Applied Biosystems SOLiD System.


548 semiconductor device enabling non-optical genome sequencing. Nature.
549 2011;475(7356):348–52.
553 Science International: Genetics Evaluation of next generation mtGenome sequencing
555 52. Ballard D. Analysis of Mitochondrial Control Region Using Sanger Sequencing. In
561 of the precision ID mtDNA whole genome panel on two massively parallel sequencing
563 55. Ovchinnikov I V., Malek MJ, Kjelland K, Drees K. Whole Human Mitochondrial DNA
565 56. Churchill JD, Stoljarova M, King JL, Budowle B. Massively parallel sequencing-enabled
567 57. Templeton JEL, Brotherton PM, Llamas B, Soubrier J, Haak W, Cooper A, et al. DNA
568 capture and next-generation sequencing can recover whole mitochondrial genomes from


82. Allard MW, Wilson MR, Monson KL, Budowle B. Control region sequences for East Asian individuals in the Scientific Working Group on DNA Analysis Methods forensic


Figure and Table legends

Figure 1 – The human mitochondrial DNA genome with genes and control regions labeled. Adapted from Shokolenko et al., 2014 (11).

Figure 2 - Phylogenetic tree of global human mitochondrial DNA variation. mtDNA tree Build 17 (18 Feb 2016). Available at http://www.phylotree.org/tree/index.htm. mtDNA-MRCA- Most recent common ancestor of mtDNA.

Figure 3 - Representation of the geographical origin of mtDNA haplogroups and main mutations that are at the origin of each haplogroup. Adapted from Kivisild et al., 2015 (83).

Table 1 - Selected published cases of human identification with mtDNA.

Table 2 - Guidelines of the DNA Commission of the International Society of Forensic Genetics, 2014.
<table>
<thead>
<tr>
<th>Table 1 (on next page)</th>
</tr>
</thead>
</table>

Selected cases of human identification with mtDNA

Selected published cases of human identification with mtDNA
<table>
<thead>
<tr>
<th>Reference/Year</th>
<th>Studied samples</th>
<th>mtDNA studied regions</th>
<th>Used methodologies</th>
<th>Reference samples</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deng YJ, Li YZ, Yu XG, Li L, Wu DY, Zhou J, et al. Preliminary DNA identification for the tsunami victims in Thailand. Genomics, Proteomics Bioinforma. 2005;3(3):143–57.</td>
<td>258 tooth samples from killed people at the 2004 Southeast Asia Thailand</td>
<td>HVI, HVII</td>
<td>PCR for amplification</td>
<td>200 relatives of the tsunami victims</td>
<td>200 tsunami victims have been identified, including both Thai nationals and foreign tourists from several nations</td>
</tr>
<tr>
<td>Reference</td>
<td>Identification process</td>
<td>PCR for amplification</td>
<td>Sanger sequencing</td>
<td>Results</td>
<td>Notes</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>-------------------------------</td>
<td>------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Rios L, Garcia-Rubio A, Martínez B, Alonso A, Puente J. Identification process in mass graves from the Spanish Civil War II. Forensic Sci Int. 2010;219(1–3).</td>
<td>Skeletal remains exhumed from a mass grave from the Spanish Civil War (1936–1939)</td>
<td>HVI, HVII</td>
<td>PCR for amplification</td>
<td>Sister of the youngest person presumptively known to be buried in the grave</td>
<td>Match between mtDNA profiles of the biologically youngest skeleton and the sister of the youngest person presumptively known to be buried in the grave</td>
</tr>
</tbody>
</table>
Table 2

Guidelines of the DNA Commission of the International Society of Forensic Genetics, 2014
### Table 2 - Guidelines of the DNA Commission of the International Society of Forensic Genetics, 2014

<table>
<thead>
<tr>
<th>Addressed recommendations/good laboratory practice</th>
<th>Recommendation</th>
<th>Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recommendation #1</td>
<td>Good laboratory practice and specific protocols for work with mtDNA must be followed in accordance with previous guidelines</td>
<td></td>
</tr>
<tr>
<td>Recommendation #2</td>
<td>Negative and positive controls as well as extraction reagent blanks must be carried through the entire laboratory process</td>
<td></td>
</tr>
<tr>
<td>Recommendation #3</td>
<td>Reported consensus sequences must be based on redundant sequence information, using forward and reverse sequencing reactions whenever practical</td>
<td></td>
</tr>
<tr>
<td>Recommendation #4</td>
<td>Manual transcription of data should be avoided and independent confirmation of consensus haplotypes by two scientists must be performed</td>
<td></td>
</tr>
<tr>
<td>Recommendation #5</td>
<td>Laboratories using mtDNA typing in forensic casework shall participate regularly in suitable proficiency testing programs</td>
<td></td>
</tr>
</tbody>
</table>

| Targeted region, amplification and sequencing ranges | Recommendation #6 | In population genetic studies for forensic databasing purposes, the entire mitochondrial DNA control region should be sequenced. |

| Reference sequence | Recommendation #7 | MtDNA sequences should be aligned and reported relative to the revised Cambridge Reference Sequence (rCRS, NC001807), and should include the interpretation range (excluding primer sequence information) |

| Alignment and notation | Recommendation #8 | IUPAC conventions using capital letters shall be used to describe differences to the rCRS and (point heteroplasmic) mixtures. Lower case letters should be used to indicate mixtures between deleted and non-deleted (inserted and non-inserted) bases. N-designations should only be used when all four bases are observed at a single position (or if no base call can be made at a given position). For the representation of deletions, “DEL”, “del” or “S” shall be used |
| Recommendation #9 | The alignment and notation of mtDNA sequences should be performed in agreement with the mitochondrial phylogeny (established patterns of mutations). Tools to assist with the notation of mtDNA sequences are available at http://empop.org/ |

| Heteroplasmy | Recommendation #10 | In forensic casework, laboratories must establish their own interpretation and reporting guidelines for observed length and point heteroplasmy. The evaluation of heteroplasmy depends on the limitations of the technology and the quality of the sequencing reactions as well as the experience of the laboratory. Differences in both PHP and LHP do not constitute evidence for excluding two otherwise identical haplotypes as deriving from the same source or same maternal lineage |
| Recommendation #11 | For population database samples, length heteroplasmy in homopolymeric sequence stretches should be interpreted by calling the dominant variant, which can be determined by identifying the position with the highest representation of a non-repetitive peak downstream of the affected stretch |

| Haplogrouping of mtDNA sequences | Recommendation #12 | MtDNA population data should be subjected to analytical software tools that facilitate phylogenetic checks for data quality control. A comprehensive suite of QC tools is provided by EMPOP |

| Databases and database searches | Recommendation #13 | The entire database of available sequences should be searched with respect to the sequencing (interpretation) range to avoid biased query results |
| Recommendation #14 | Laboratories must be able to justify the choice of database(s) and statistical approach used in reporting |
| Recommendation #15 | Laboratories must establish statistical guidelines for use in reporting an mtDNA match between two samples |
| Recommendation #16 | Highly variable positions such as length variants in homopolymeric stretches should be disregarded from searches for determining frequency estimates. Heteroplasmic calls should be queried in a manner that |
does not exclude any of the heteroplasmic variants
Figure 1 (on next page)

Genes and control regions labeled of the human mitochondrial DNA genome

The human mitochondrial DNA genome with genes and control regions labeled. Adapted from Shokolenko et al., 2014 (11)
Figure 2 (on next page)

Phylogenetic tree of global human mitochondrial DNA variation

mtDNA-MRCA
Figure 3 (on next page)

Geographical origin of mtDNA haplogroups and main mutations that are at the origin of each haplogroup

Representation of the geographical origin of mtDNA haplogroups and main mutations that are at the origin of each haplogroup. Adapted from Kivisild et al., 2015 (83).