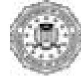
**U.S. Department of Justice**
Federal Bureau of Investigation





FORENSIC SCIENCE COMMUNICATIONS
July 1999 Volume 1 Number 2

# Mitochondrial DNA Analysis at the FBI Laboratory

Alice R. Isenberg
Forensic Examiner/Biologist

Jodi M. Moore
Biologist

DNA Analysis Unit II
Federal Bureau of Investigation
Washington, DC

## Introduction

The FBI Laboratory began conducting studies on the feasibility of mitochondrial DNA (mtDNA) analysis for human identity testing in the late 1980s. Laboratory research began on a protocol for using mtDNA sequencing in forensic casework in 1992. After the sequencing technique was validated, examinations on evidentiary samples began in June 1996.

MtDNA sequencing is often used in cases where biological evidence may be degraded or small in quantity. Cases in which hairs, bones, or teeth are the only evidence retrieved from a crime scene are particularly well-suited to mtDNA analysis. Missing persons cases can benefit from mtDNA testing when skeletonized remains are recovered and compared to samples from the maternal relatives or personal effects of missing individuals. Also, hairs recovered at crime scenes can often be used to include or exclude individuals using mtDNA testing. This review will examine the process of mitochondrial DNA typing, including the interpretation of results, the phenomenon of heteroplasmy, the mtDNA population database, presentation of mtDNA population statistics, quality assurance issues, and testimonial experience.

*Back to the topical list*

## Mitochondrial DNA Background

Mitochondrial DNA differs from nuclear DNA in its location, its sequence, its quantity in the cell, and its mode of inheritance (Figure 1). The nucleus of the cell contains two sets of 23 chromosomes—one paternal set and one maternal set. However, cells may contain hundreds to thousands of mitochondria, each of which may contain several copies of mtDNA. Nuclear DNA has many more bases than mtDNA, but mtDNA is present in many more copies than nuclear DNA. This characteristic of mtDNA is

useful in situations where the amount of DNA in a sample is very limited. Typical sources of DNA recovered from crime scenes include hair, bones, teeth, and body fluids such as saliva, semen, and blood.

In humans, mitochondrial DNA is inherited strictly from the mother (Case and Wallace 1981; Giles et al. 1980; Hutchison et al. 1974). Thus, the mtDNA sequences obtained from maternally related individuals, such as a brother and a sister or a mother and a daughter, will exactly match each other in the absence of a mutation (Figure 2). This characteristic of mtDNA is advantageous in missing persons cases as reference mtDNA samples can be supplied by any maternal relative of the missing individual (Ginther et al. 1992; Holland et al. 1993; Stoneking et al. 1991). However, mtDNA analysis is limited when compared to nuclear DNA analysis in that it cannot discriminate between individuals of the same maternal lineage.

The human mtDNA genome is approximately 16,569 bases in length and has two general regions: the coding region and the control region (Figure 3). The coding region is responsible for the production of various biological molecules involved in the process of energy production in the cell. The control region is responsible for regulation of the mtDNA molecule. Two regions of mtDNA within the control region have been found to be highly polymorphic, or variable, within the human population (Greenberg et al. 1983). These two regions are termed Hypervariable Region I (HV1), which has an approximate length of 342 base pairs (bp), and Hypervariable Region II (HV2), which has an approximate length of 268 bp. Forensic mtDNA examinations are performed using these two regions because of the high degree of variability found among individuals.

Approximately 610 bp of mtDNA are currently sequenced in forensic mtDNA analysis. Recording and comparing mtDNA sequences would be difficult and potentially confusing if all of the bases were listed. Thus, mtDNA sequence information is recorded by listing only the differences with respect to a reference DNA sequence. By convention, human mtDNA sequences are described using the first complete published mtDNA sequence as a reference (Anderson et al. 1981). This sequence is commonly referred to as the Anderson sequence. It is also called the Cambridge reference sequence or the Oxford sequence. Each base pair in this sequence is assigned a number. Deviations from this reference sequence are recorded as the number of the position demonstrating a difference and a letter designation of the different base. For example, a transition from A to G at Position 263 would be recorded as 263 G. If deletions or insertions of bases are present in the mtDNA, these differences are denoted as well.

*Back to the topical list*

## Analysis Procedures

Presently, the forensic analysis of mtDNA is rigorous and labor-intensive. Several molecular biological techniques are combined to obtain a mtDNA sequence from a sample. The steps of the mtDNA analysis process include primary visual analysis, sample preparation, DNA extraction, polymerase chain reaction (PCR) amplification, postamplification

quantification of the DNA, automated DNA sequencing, and data analysis.

### Step 1: Primary Visual Analysis

The first step in the analysis of hair is a microscopic comparison of evidentiary hair and a sample population of reference hairs. The sample is mounted on a glass microscope slide and viewed under a comparison microscope to observe microscopic characteristics of the sample compared to known hair standards (Figure 4). If the hair does not exhibit the same microscopic characteristics as the known standard, mtDNA analysis generally is not performed. If the hair from a questioned source exhibits similar microscopic characteristics as hair from a known source, however, mtDNA analysis is performed in order to determine on a molecular level if the hair is consistent with reference standards from a particular individual. At times, hairs are not microscopically compared prior to mtDNA analysis because they are not suitable for microscopic comparison. Hairs that are not suitable for significant comparison purposes are screened microscopically, and limited information is assessed as a basis for possible exclusion. If the hairs cannot be excluded following this limited examination, mtDNA analysis may be performed.

In instances in which bone or tooth material are collected, the tissue is first inspected by an anthropologist or forensic odontologist. If the tissue is of human origin, mtDNA analysis can be used in conjunction with medical, anthropological, and odontological examinations to assist in the identification process.

*Back to the topical list*

### Step 2: Sample Preparation

Evidentiary samples are cleaned prior to the mtDNA sequencing process to remove contaminating materials surrounding or adhering to the sample. This step is important to ensure that the sequence of the DNA obtained from the sample originates from the sample and not from exogenous human DNA.

For hair samples, the cleaning step consists of a detergent treatment in an ultrasonic water bath. This process removes possible contaminating residues from the hair. The hair sample is then placed in an extraction solution and ground using a small mortar and pestle. The DNA is released from the cellular material, resulting in a homogenate that contains both the cellular material and the released DNA.

Bone and tooth samples also undergo a cleaning process. To clean a bone or tooth, the exterior is sanded to remove any extraneous material that may be adhering to the surface. A small sample of the tissue is then removed and ground into a fine powder. A tooth is cross-sectioned, and the dentin and the pulp (the inner layers) are used for DNA extraction. In fresh teeth, the DNA is generally extracted from the fleshy pulp. In extremely old teeth, the pulp dehydrates and the DNA may be extracted from the compact dentin layer. The powdered bone and teeth are also placed in a solution to release the DNA from the cells.

*Back to the topical list*

### Step 3: DNA Extraction

To extract DNA, the cellular homogenate from the sample preparation step is exposed to a mixture of organic chemicals that separate the DNA from other biological molecules, such as proteins. The mixture is spun in a centrifuge, and the DNA remains soluble in the top water-based layer. The rest of the cellular components are soluble in the bottom organic layer or in the interface between the two layers. The top layer is filtered and concentrated. The DNA sample is now purified and ready for the PCR amplification process.

### Step 4: Amplification by the Polymerase Chain Reaction

The polymerase chain reaction (PCR) is a procedure that makes many copies of a small amount of DNA (Figure 5). In the first step of the PCR process, the two strands of the DNA double helix are separated by heating the sample. These two separated strands are called the template strands. A new DNA strand is then made from each template using an enzyme that copies the existing DNA molecule. This copying process is repeated a number of times, and during each repetitive cycle, the amount of DNA in the reaction tube is theoretically doubled. At the end of this process, many millions of copies of the original targeted DNA in the extract are present.

*Back to the topical list*

### Step 5: Postamplification Purification and Quantification

The DNA generated by the PCR is purified and quantified prior to the sequencing process. The purification is performed using filtration devices that remove the excess reagents used in the PCR from the sample. The quantification step is performed using capillary electrophoresis (CE). This technique compares the amount of DNA in the PCR product to a known DNA standard to determine the concentration of the DNA in the PCR-amplified sample (Figure 6). Amplified and quantified blank samples (negative controls) and known DNAs (positive controls) are also included to indicate whether any exogenous DNA is present and whether the amplification was successful. If the PCR failed to yield product, the sample may be reextracted and reamplified. If the PCR is successful, the products are prepared for DNA sequencing.

### Step 6: Sequencing

The dideoxy terminator method, also called Sanger's Method (Sanger et al. 1977), is used for mtDNA cycle sequencing. The cycle sequencing process (Carothers et al. 1989; Murray 1989) is similar to the PCR, but slightly different chemicals are used (Figure 7). A set of terminator bases are used in addition to the normal bases that elongate the growing strand of DNA. These terminator bases lack a chemical group that would normally allow the enzyme to place another base after them. The altered bases also carry a fluorescent dye that is readily detected by an automated instrument. The normal bases compete with the altered bases for incorporation into the growing DNA strand, resulting in a collection of DNA products that differ in size by one base and have a fluorescently labeled base at the end position.

The products resulting from the sequencing reaction are separated on the basis of their length using a technique called gel electrophoresis. The size of the pores in a gel electrophoresis matrix regulate the rate at which each DNA fragment travels through the gel under the influence of an electric

field. Smaller products travel more quickly through the pores, whereas the longer products are retarded by the constricting pore size. The labeled DNA fragments begin from the same point on the gel, and the fluorescence detector records the emitted wavelength of the fluorescent dyes on each base as the fragments travel past the detection area of the instrument. The instrument generates a chromatogram, or colored graph, depicting the colors of the labeled fragments one base at a time (Figure 8). The sequence of the mtDNA is determined from a series of cycle sequencing reactions.

*Back to the topical list*

## Interpretation Guidelines

Mitochondrial DNA sequences are initially generated by computer software and subsequently edited by a DNA examiner using multiple runs from the same template to obtain the final sequence of the DNA. After editing, the sequence is recorded by listing the difference or differences from the Anderson reference sequence. The sequences obtained from the samples of questioned origin are compared to the sequences of the known, or reference, samples to determine if any points of difference are present between the samples (Figure 9). The FBI Laboratory has established guidelines for the interpretation of the differences and similarities between samples. Basically, samples cannot be excluded as originating from the same source if there is a sequence concordance. A sequence concordance is the presence of the same base (or a common base) at every position analyzed.

The human body contains trillions of cells, each of which can contain thousands of copies of the mtDNA genome. Complete homoplasmy (the same sequence of mtDNA) for each of these mtDNA molecules would be surprising because of the immense amounts of mtDNA present in the body. Thus, heteroplasmy is expected to be present at some level in all individuals. Heteroplasmy is the occurrence of more than one type at a particular position in a DNA sequence, and there are two forms of heteroplasmy found in mtDNA. Sequence heteroplasmy, or point heteroplasmy, is the occurrence of more than one base at a particular position or positions in the mtDNA sequence. Length heteroplasmy is the occurrence of more than one length of a stretch of the same base in a mtDNA sequence. The detection methods currently available to molecular biologists cannot detect low levels of heteroplasmy. For this reason, heteroplasmy is an operational term used when the current scientific methods are capable of detecting more than one sequence in an individual.

Heteroplasmy was first observed in forensic mtDNA sequences in 1994 by the Forensic Science Service (FSS) in the United Kingdom while identifying the remains of the Romanov family (Gill et al. 1994). The phenomenon of heteroplasmy was studied in a controlled setting by the FBI Laboratory (Wilson et al. 1997). This study led to a better understanding of the manner in which heteroplasmy exhibits itself in forensic samples. The combination of recent improvements in sequencing enzymes and fluorescent dye chemistry has enhanced the heteroplasmy-detection capabilities of the forensic community. When heteroplasmy is observed at the same position in a questioned and a known sample and all

of the other bases are the same, the significance of the match is enhanced.

The level of heteroplasmy may not always be the same in various tissues. For example, a hair may contain mostly C at a particular position in the mtDNA genome, but a blood sample from the same individual may contain equal amounts of C and T at the same position. This phenomenon is due to the different mechanisms by which cells are generated in different tissues. If different tissues demonstrate heteroplasmy with the presence of common bases at every position, then a sequence concordance is present, and one cannot exclude two samples as originally coming from the same source or maternal lineage. In cases where heteroplasmy is thought to occur, additional known samples can be sequenced to determine if the heteroplasmy is visible in other tissues. Obviously, further testing cannot always be performed on a crime scene sample of limited quantity, but it can prove helpful for interpretation of known samples.

*Back to the topical list*

## Population Database

The FBI Laboratory, the Armed Forces DNA Identification Laboratory, and other laboratories have collaborated to compile a mtDNA population database containing the sequences from Hypervariable Regions I and II. The database is referred to as the SWGDAM (Scientific Working Group on DNA Analysis Methods) database. It contains sequences from four main racial groups: Caucasians, Africans, Hispanics, and Asians. Most of these samples have been obtained from paternity-testing laboratories, blood banks, or academic groups studying ethnic populations. The database currently contains 2,426 mtDNA sequences from unrelated individuals. However, the database is updated frequently and is constantly growing. Parts of the control region outside the two hypervariable segments are also being analyzed for enhanced discrimination potential.

When a sequence from a questioned sample and a known sample is the same, the SWGDAM database is searched for this sequence. Any indeterminate bases are searched as Ns, where N could be A, C, G, or T, to account for the possibility that any of the four bases might occur at that position. The FBI Laboratory lists the number of observations of a sequence in each racial subgroup of the database in a report of a mtDNA examination. For example, a sequence might be seen five times in the database samples of Caucasian descent and one time in the database samples of Hispanic descent yet not appear in the remaining database subgroups.

*Back to the topical list*

## Reporting Statistics

Mitochondrial DNA is a single genetic locus. The different regions of mtDNA are not independent of each other because they are inherited together with the entire mtDNA genome. Most of the sequences in the forensic mtDNA database occur a single time (approximately 60 percent), and the total number of mtDNA sequences in the entire human population

is not known. Reliable frequency estimates for most mtDNA sequences are therefore not possible. The mtDNA database is too small at this time to provide estimates of the frequency of occurrence of most mtDNA sequences in the population. This limitation can be illustrated using a simple analogy. If a jury of 12 people was used as a database to estimate the occurrence of brown hair in the general population of the United States, the number of individuals in the jury with brown hair might closely reflect the frequency of brown hair in the general population because brown hair is common. If the same database were used to estimate the number of people who are 7 feet tall, the frequency would not be accurate. The jury database would probably not contain any individuals who were 7 feet tall. Thus, the estimated frequency for this event would be zero. However, it is known that very tall individuals do exist at some low frequency in the population. If one used the jury as a database and conservatively estimated the frequency of people who are 7 feet tall to be approximately 1 in 12, this estimate would obviously be incorrect. The jury database is an extreme example, but it illustrates the fact that small databases are not effective tools for estimating frequencies of rare events.

Currently, the FBI Laboratory does not provide frequency estimates of mtDNA types in laboratory reports because of the restrictions mentioned previously involving the database size. The FBI states only the number of occurrences of a mtDNA sequence in the current database. To date, the number of mtDNA sequences present in the general population is unknown because not all sequences have yet been observed. However, statistical methods exist for calculating an upper-bound estimate of the frequency of mtDNA types with zero occurrences or very few occurrences in a database of limited size. This upper-bound estimate describes the highest frequency expected for a particular mtDNA sequence using the database. The entire range of probable frequencies, between upper and lower bounds, can likewise be provided. As the database grows in size, the frequency estimates for individual mtDNA profiles will become more and more refined and eventually lead to reliable population frequency estimates.

*Back to the topical list*

## Quality Assurance

Many steps are taken during the mtDNA sequencing process to ensure the quality of the results (see *FBI Laboratory DNA Unit II Mitochondrial DNA Sequencing Protocol*, Revised May 1998). The FBI Laboratory is accredited by the American Society of Crime Laboratory Directors/Laboratory Accreditation Board (ASCLD/LAB), and the DNA Analysis Units follow the quality assurance guidelines established by the Scientific Working Group on DNA Analysis Methods (SWGDAM), the DNA Advisory Board (DAB), and ASCLD/LAB. Two open and external proficiency tests are conducted annually for all unit personnel working cases, and a record of these tests is retained. General and specific precautions designed to minimize contamination are followed, including the use of gloves, laboratory coats, laminar flow or deadspace hoods, and aerosol-resistant pipet tips. During an analytical procedure, only one item of evidence from a case is opened at a time. Items of questioned origin are analyzed prior to items of known origin in order to minimize the potential of carryover from knowns into questioned items. Preamplification areas are

physically separated from postamplification areas to prevent amplified DNA product transfer into unamplified samples. Most reagents, workspaces, and instruments such as pipets are subjected to isopropanol washings, 10-percent bleach washings, ultraviolet (UV) irradiation, or combinations of these treatments. These steps are taken in an attempt to remove and destroy extraneous DNA molecules prior to the extraction and amplification steps. Also, all instruments used in the examination process are routinely calibrated. The instruments subjected to quality screening, monitoring, or both include the thermal cycler used for PCR and sequencing, the capillary electrophoresis instrument, the sequencing instrument, the water baths, all thermometers, freezers and refrigerators, and the pipets used for delivering reagents into reaction mixtures.

Reagent blanks and negative controls are prepared containing all chemicals used in the extraction and amplification process, respectively. The reagent blank and negative control samples go through all of the same procedural steps as evidentiary samples. These controls are used to determine the presence and amount of exogenous DNA in the reagents and equipment. Positive controls are used in the amplification and sequencing steps to monitor the success of amplification and sequencing. The mtDNA analysis of a sample does not proceed if the positive control fails to amplify or provide a quality sequence or both.

Capillary electrophoresis with laser-induced fluorescence detection is used to determine the DNA concentration of samples following amplification. The mtDNA analysis of a sample will not proceed if the amount of amplified DNA in the reagent blank or negative control exceeds 10 percent of the amount of amplified DNA in the questioned sample. Samples are reextracted with new reagents (and a new reagent blank) should the amount of amplified DNA in the reagent blank or negative control exceed the 10-percent threshold. The 10-percent threshold was established in an early FBI study (Wilson et al. 1995), and the correct sequences for hundreds of samples have been obtained in controlled studies since the inception of the threshold. Moreover, should the DNA sequence of a reagent blank or negative control be the same as that of a sample, but less than 10 percent of the quantity of the sample, the analysis is repeated or omitted. If typeable mtDNA sequence is obtained from the reagent blank, an attempt is made to determine the source of this DNA. Mitochondrial DNA sequences from all laboratory personnel analyzing mtDNA or handling mtDNA evidence are kept on file for this purpose. Because of the sensitivity of the mtDNA amplification process, even minute amounts of DNA present in reagent blanks and negative controls can amplify to a detectable level. Most of this DNA is not of sufficient quality to be analyzed, but it is examined independently by two DNA examiners before this determination is made. Finally, when sequencing reactions are performed, the mtDNA is sequenced in two directions, using both strands of the mtDNA molecule. In cases of sequence concordance, each sequence obtained is independently analyzed by at least two examiners, and all confirmations are recorded.

*Back to the topical list*

## Testimony

In the United States, there are seven laboratories currently conducting

forensic mtDNA examinations: the FBI Laboratory; Laboratory Corporation of America (LabCorp) in Research Triangle Park, North Carolina; Mitotyping Technologies in State College, Pennsylvania; the Bode Technology Group (BTG) in Springfield, Virginia; the Armed Forces DNA Identification Laboratory (AFDIL) in Rockville, Maryland; BioSynthesis, Inc. in Lewisville, Texas; and Reliagene in New Orleans, Louisiana.

Mitochondrial DNA analyses have been admitted in criminal proceedings from these laboratories in the following states as of April 1999: Alabama, Arkansas, Florida, Indiana, Illinois, Maryland, Michigan, New Mexico, North Carolina, Pennsylvania, South Carolina, Tennessee, Texas, and Washington. Mitochondrial DNA has also been admitted and used in criminal trials in Australia, the United Kingdom, and several other European countries.

*Back to the topical list*

## Concluding Remarks

Since 1996, the number of individuals performing mitochondrial DNA analysis at the FBI Laboratory has grown from 4 to 12, with more personnel expected in the near future. Over 150 mitochondrial DNA cases have been completed by the FBI Laboratory as of March 1999, and dozens more await analysis. Forensic courses are being taught by the FBI Laboratory personnel and other groups to educate forensic scientists in the procedures and interpretation of mtDNA sequencing. More and more individuals are learning about the value of mtDNA sequencing for obtaining useful information from evidentiary samples that are small, degraded, or both. Mitochondrial DNA sequencing is becoming known not only as an exclusionary tool but also as a complementary technique for use with other human identification procedures. Mitochondrial DNA analysis will continue to be a powerful tool for law enforcement officials in the years to come as other applications are developed, validated, and applied to forensic evidence.

*Back to the topical list*

## References

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R., and Young, I. G. Sequence and organization of the human mitochondrial genome, *Nature* (1981) 290: 457–465.

Carothers, A. M., Urlaub, G., Mucha, J., Grunberger, D., and Chasin, L. A. Point mutation analysis in a mammalian gene: Rapid preparation of total RNA, PCR amplification of cDNA, and Taq sequencing by a novel method, *BioTechniques* (1989) 7:494–499.

Case, J. T. and Wallace, D. C. Maternal inheritance of mitochondrial DNA polymorphisms in cultured human fibroblasts, *Somatic Cell Genetics*

(1981) 7:103–108.

Giles, R. E., Blanc, H., Cann, H. M., and Wallace, D. C. Maternal inheritance of human mitochondrial DNA, *Proceedings of the National Academy of Sciences* (1980) 77:6715–6719.

Gill, P., Ivanov, P. L., Kimpton, C., Piercy, R., Benson, N., Tully, G., Evett, I., Hagelberg, E., and Sullivan, K. Identification of the remains of the Romanov family by DNA analysis, *Nature Genetics* (1994) 6:130–135.

Ginther, C., Issel-Tarver, L., and King, M. C. Identifying individuals by sequencing mitochondrial DNA from teeth, *Nature Genetics* (1992) 2:135–138.

Greenberg, B. D., Newbold, J. E., and Sugino, A. Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA, *Gene* (1983) 21:33–49.

Holland, M. M., Fisher, D. L., Mitchell, L. G., Rodriquez, W. C., Canik, J. J., Merril, C. R., and Weedn, V. W. Mitochondrial DNA sequence analysis of human skeletal remains: Identification of remains from the Vietnam War, *Journal of Forensic Sciences* (1993) 38:542–553.

Hutchison, C. A., Newbold, J. E., Potter, S. S., and Edgell, M. H. Maternal inheritance of mammalian mitochondrial DNA, *Nature* (1974) 251:536–538.

Murray, V. Improved double-stranded DNA sequencing using the linear polymerase chain reaction, *Nucleic Acids Research* (1989) 17:8889.

Sanger, F., Nicklen, S., and Coulson, A. R. DNA sequencing with chain-terminating inhibitors, *Proceedings of the National Academy of Sciences* (1977) 74: 5463–5467.

Stoneking, M., Hedgecock, D., Higuchi, R. G., Vigilant, L., and Erlich, H. A. Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes, *American Journal of Human Genetics* (1991) 48:370–382.

Wilson, M. R., Polanskey, D., Butler, J., DiZinno, J. A., Repogle, J., and Budowle, B. Extraction, PCR amplification, and sequencing of mitochondrial DNA from human hair shafts, *BioTechniques* (1995) 18:662–669.

Wilson, M. R., Polanskey, D., Repogle, J., DiZinno, J. A., and Budowle, B. A family exhibiting heteroplasmy in the human mitochondrial DNA control region reveals both somatic mosaicism and pronounced segregation of mitotypes, *Human Genetics* (1997) 100:167–171.