MITOCHONDRIAL DNA: TODAY AND TOMORROW

Constance L. Fisher, Alice R. Isenberg, John E. B. Stewart, Kevin W. P. Miller, Catherine E. Theisen, Mark R. Wilson, Joseph A. DiZinno, Bruce Budowle
Laboratory Division, Federal Bureau of Investigation

Mitochondrial DNA (mtDNA) analysis has proven to be a powerful exclusionary tool in forensic casework (1). MtDNA analysis is particularly useful in cases involving evidence such as hair shafts and skeletal remains. For a number of reasons, one can obtain a mtDNA sequence in instances where a nuclear DNA (nDNA) result could not be obtained. As more and more investigators have learned about the uses of mtDNA analysis, the number of submitted cases requesting mtDNA analysis has increased.

Some of the notable features that differ between nDNA and mtDNA include the mode of inheritance, copy number, and degree of polymorphism. MtDNA is maternally inherited and treated as a single locus due to the lack of recombination. Thus, excluding mutation, all maternal relatives will have the same mtDNA sequence. Because of this, mtDNA can never be used to uniquely identify an individual. However, this maternal relatedness feature of mtDNA genetics can be advantageous since maternal relatives can serve as known reference sources to aid in the identification of unknown human remains. Another feature of mtDNA and nDNA involves copy number. There are two copies of the nDNA genome per cell. With mtDNA, several copies of the mtDNA genome are found within each mitochondrion. Since many mitochondria are found within the cell, on average, hundreds to thousands of copies of the mtDNA genome can be found within a single cell. Because of this increase in copy number, mtDNA analysis is more likely to be successful on samples with very low levels of nDNA. A third important feature is that mtDNA is highly polymorphic. To date, over 2500 different mtDNA sequences have been seen examining only the HV1 and HV2 regions of the mtDNA genome in approximately 4000 individuals. In contrast, even the highly polymorphic nDNA loci used in RFLP analysis have fewer than 100 operationally defined alleles.

Not surprisingly, as the demand has increased for mtDNA analysis, the unit in the FBI laboratory that conducts mtDNA casework analysis has grown. Initially, only four people were assigned to the unit in 1996 when the FBI first started accepting mtDNA casework. In 2000, thirteen individuals are involved in processing mtDNA casework, with an additional four individuals assigned specifically to running the National Missing Persons DNA Database Program (discussed below). More than 300 mtDNA cases have been processed by the FBI Laboratory, involving the analysis of over 700 items. About two-thirds of the questioned items analyzed have been hairs, with the remainder being bones and teeth. Microscopic comparison of head and pubic hairs is used as a screening tool prior to mtDNA analysis. About 90% of hairs associated with an individual by microscopic analysis also have the same mtDNA sequence as the associated individual. All hairs that have been excluded by microscopic comparison have also been excluded by mtDNA analysis, although it is anticipated that a microscopic exclusion of a maternal relative will result in a failure to exclude by mtDNA analysis.

MtDNA analysis has been accepted in a number of court proceedings. At this date, mtDNA analysis has been accepted in Federal court as well as in at least 26 state courts, including Alabama, Arkansas, California, Connecticut, Florida, Georgia, Hawaii, Illinois, Indiana, Kansas, Louisiana, Maryland, Michigan, Minnesota, New Mexico, New York, North Carolina, Ohio, Pennsylvania, South Carolina, Texas, Tennessee, Vermont, Virginia, Washington, and West Virginia. Lately, the number of requests for admissibility hearings has diminished substantially.

Some common issues that have been raised by experts in cases involving mtDNA analysis include contamination, paternal leakage and recombination, heteroplasmy, and statistical inferences with the mtDNA population database. Contamination is a concern with all PCR-based techniques since the presence of exogenous DNA could lead to confusing results. Typically, over 30 amplification cycles are required for mtDNA typing. As the number of cycles increases, the opportunity for detectable contamination increases. However, numerous measures are undertaken to minimize contamination, which include: individuals that perform the analysis wear protective clothing (lab coats, gloves, safety glasses, and masks); all tubes and appropriate reagents are sterilized and exposed to ultraviolet light; all extractions are performed inside hoods, which have been cleaned with bleach and exposed to ultraviolet light; questioned samples are processed before known samples, so that known samples cannot be a source of contamination for questioned items; separate equipment is dedicated for working with questioned and known samples; and extraction and amplification set-up processes are conducted in separate rooms from amplification and sequencing. In addition to the measures taken to minimize contamination, reagent blank and positive and negative amplification controls are prepared to assess and monitor the level of contamination. Mixture experiments have shown that if the contaminating DNA is less than 10% of the questioned or known sample, it will not affect the interpretation of the resulting sequence (2). Nevertheless, reagent blanks and negative controls which yield a typeable sequence that is the same as the sample will result in either re-extraction or not using that region of the sequence for comparision, even if the 10% threshold is not violated. These efforts are usually sufficient to demonstrate that contamination did not affect the interpreted sequences obtained in the case.

Another issue frequently raised in court involves paternal leakage and recombination. Hagelberg, *et al.* (3) reported that an island population was found a to have a rare mutation in a number of distinct lineages. These authors suggested that paternal leakage and recombination was the most likely explanation for this observation (3). However, it was later reported that a misalignment of sequences produced the apparent rare mutation and the proposed mechanism of paternal leakage and recombination is no longer viable (4). Awadalla, *et al*. (5) found a decline in linkage disequilibrium as a function of distance between polymorphic sites when examining hominid mtDNA sequences, and attributed this observation to recombination. However, subsequent studies have found no evidence for recombination when examining other mtDNA data sets (6). Additionally, no empirical mother and child mtDNA inheritance studies have shown evidence of paternal leakage or recombination.

Another issue frequently raised in cases involving mtDNA analysis is heteroplasmy. Heteroplasmy occurs when an individual has more than one mtDNA sequence and can be manifested in two ways: sequence and length. Sequence, or point, heteroplasmy arises when the two sequences differ by a base substitution, resulting in an apparent mixture at the heteroplasmic

site. Length heteroplasmy arises when the two sequences differ in the number of bases in a homopolymeric region and yields a characteristic out-of-phase pattern downstream from the point of heteroplasmy. Sequence heteroplasmy can theoretically occur anywhere in the mtDNA genome, although hotspot positions have been observed. Length heteroplasmy occurs most often in the homopolymeric stretches in HV1 (16184-16193) and HV2 (303-310). Generally, the majority of samples analyzed are operationally defined as homoplasmic. In samples from individuals who demonstrate heteroplasmy, it generally occurs only at one location within the individual's mtDNA, resulting in sequences that differ only at that location. Numerous studies have demonstrated that the level of heteroplasmy can vary considerably within an individual as well as between generations (7-15). Nevertheless, if two samples demonstrate sequence concordance, or a common base at every position, they cannot be excluded as coming from the same source.

The SWGDAM (Scientific Working Group on DNA Analysis Methods) mtDNA population database is used by forensic laboratories to provide an estimate of the relative rarity of mtDNA sequences. The database is a compilation of sequences obtained from several laboratories and is updated regularly. The database is currently searched using MitoSearch, an FBI-developed software package that allows for identification of sequences and counting of the number of observations of a particular sequence in the database. The latest version of MitoSearch has the capability of searching the entire mtDNA control region. CODIS^{mt} is under development and will have the benefits of national searching capability as well as a help desk and technical support. The indices in CODIS^{mt} will include: forensic unknown, forensic known, unidentified persons, and unidentified persons reference. The mtDNA population database will be accessible in a manner similar to Popstats used for nDNA. CODIS^{mt} will be available for release in January 2001.

Prior to the release of the July 2000 version of the mtDNA population database, the sequences were subjected to a quality assurance review. The resulting forensic section of the database contains mtDNA sequences from 4142 individuals composed of: 1773 Caucasians, 794 African-Americans, 110 Sierra Leoneans, 24 Egyptians, 694 Hispanics, 163 Japanese, 182 Koreans, 70 Thai, 152 Navajos, and 180 Apaches. About 2000 of these sequences contain data from the entire control region. A public section has been added to the database that includes over 3000 mtDNA sequences reported in the published literature.

Statistical analysis of the 1773 Caucasian sequences in the July 2000 version of the mtDNA population database reveals that the most common sequence in this population is seen in 7.4% of these individuals (ignoring length differences at position 309). Of the 1175 different mtDNA profiles seen in Caucasians, 55% occur only once. In contrast, the most common sequence in the African-Americans in the July 2000 version of the database is seen in only 1.1% of the 794 individuals sequenced. Of the 602 different African-American mtDNA profiles, 63% occur only once. Comparison of the overall July 2000 database with a version used in 1998 shows that the number of sequences occurring once is decreasing, going from 63% in 1998 to 54% in 2000. The percentage of single occurrences will continue to decrease as more mtDNA samples are typed. However, the number of individuals who must be sequenced to reach the limit of mtDNA diversity is unknown.

The FBI's National Missing Persons DNA Database Program is scheduled to go online in January 2001. This program will generate DNA profiles from unidentified remains as well as references from family members of persons reported missing in a criminal matter. These DNA profiles will be entered in a separate index in CODIS^{mt}. Once the program begins to solicit cases nationwide, a large backlog is anticipated since over 150 unidentified remains are discovered each year. California alone has over 2000 undiscovered remains to be analyzed, and current legislation has provided funding for the state lab system to process these remains. The FBI has successfully completed a pilot program and is currently conducting validation studies directed at optimizing protocols and testing new instrumentation.

The next approach in mtDNA testing will likely involve single nucleotide polymorphisms (SNPs). SNP analysis is similar to sequence-specific oligonucleotide testing (16). Parsimony analysis of the current SWGDAM mtDNA population database has found approximately sixty informative SNP sites using just HV1 and HV2 sequence data in the general population. It is expected that additional informative sites will be found upon analysis of specific subpopulations. Also, more informative sites will be identified when the analysis is expanded to include the entire control region, as well as the entire mtDNA genome. It is envisioned that mtDNA SNP analysis will be helpful for field use at mass disasters, and as a screening tool by laboratories not equipped to perform full mtDNA sequencing. SNP analysis should also provide additional information to distinguish the most common sequence types.

This is publication 01-05 of the Laboratory Division of the Federal Bureau of Investigation. Names of commercial manufacturers are provided for identification only, and inclusion does not imply endorsement by the Federal Bureau of Investigation.

1. Holland MM, Parsons TJ. Mitochondrial DNA sequence analysis – validation and use for forensic casework. For Sci Rev 1999; 11: 21-51.

2. Wilson MR, Polanskey D, Butler J, DiZinno JA, Replogle J, and Budowle B. Extraction, PCR amplification and sequencing of mitochondrial DNA from human hair shafts. BioTech 1995; 18: 662-669.

3. Hagelberg E, Goldman N, Liu P, Whelan S, Schiefenholven W, Clegg JB, Bowden DK. Evidence for mitochondrial DNA recombination in a human population of island Melanesia. Proc R Soc Lond 1999; 266: 485-492.

4. Hagelberg E, Goldman N, Liu P, Whelan S, Schiefenholven W, Clegg JB, Bowden DK. Evidence for mitochondrial DNA recombination in a human population of island Melanesia: correction. Proc R Soc Lond 2000; 267: 1595-1596.

5. Awadalla P, Eyre-Walker A, Smith JM. Linkage disequilibrium and recombination in hominid mitochondrial DNA. Science 1999; 286: 2524-2525.

6. Morris AAM, Lightowlers RN. Mitochondrial DNA recombination. Lancet 2000; 356:941.

7. Calloway CD, Reynolds RL, Herrin GL Jr, Anderson WW. The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. Am J Hum Genet 2000; 66: 1384-1397.

8. Bendall KE, Sykes BC. Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. Am J. Hum Genet. 1995; 57: 248-256.

9. Bendall KE, Macaulay, VA, Sykes BC. Variable levels of a heteroplasmic point mutation in individual hair roots. Am J Hum Genet 1997; 61: 1303-1308

10. Huhne J, Pfeiffer H, Brinkmann B. Heteroplasmic substitutions in the mitochondrial DNA control region in mother and child samples. Int J Legal Med 1998; 112: 27-30.

11. Lutz S., Weisser HJ, Heizmann J, Pollak, S. Mitochondrial heteroplasmy among maternally related individuals. Int J. Legal Med 2000; 113:155-161.

12. Marchington DRE, Hartshorne GM, Barlow D, Poulton J. Homopolymeric tract heteroplasmy in mtDNA from tissues and single oocytes: support for a genetic bottleneck. Am J Hum Genet 1997; 60: 408-416.

13. Stewart JEB, Fisher CL, Aagaard PJ, Wilson MR, Isenberg AR, Polanskey D, Pokorak E, DiZinno JA, Budowle B. Length variation in HV2 of the human mitochondrial DNA control region. J For Sci (in press)

14. Tully LA, Parsons TJ, Steighner RJ, Holland MM, Marino MA, Prenger VL. A sensitive denaturing gradient-gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region 1 of the human mtDNA control region. Am J Hum Genet 2000; 67: 432-443.

15. Wilson MR, Polanskey D, Replogle J, DiZinno JA, Budowle B. A family exhibiting heteroplasmy in the human mitochondrial DNA control region reveals both somatic mosaicism and pronounced segregation of mitotypes. Hum Genet 1997; 100: 167-171.

16. Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Ehrlich HA. Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence specific oligonucleotide probes. Am J Hem Genet 1991; 48: 370-382.