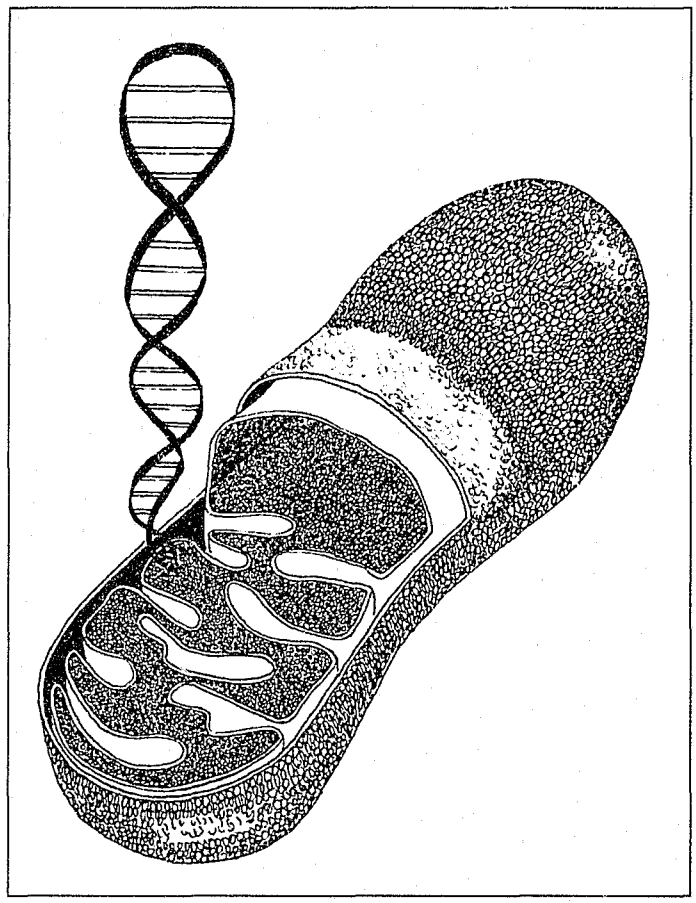
U.S. Department of Justice
Federal Bureau of Investigation

CRIME LABORATORY DIGEST

Volume 20, Number 4 October 1993



IN THIS ISSUE

CRIME LABORATORY DIGEST

Volume 20, Number 4 October 1993

ARTICLES

DEPARTMENTS

Special thanks to Glenda H. Wilson for providing her original artwork, which is depicted on the cover of this issue.

The *Crime Laboratory Digest* (ISSN 0743-1872) is a quarterly forensic science journal published by the FBI Laboratory Division in cooperation with the American Society of Crime Laboratory Directors (ASCLD). It is intended to serve as a means of communication between crime laboratories, permitting information of interest and value to be disseminated among crime laboratory scientists.

Postmaster: Send address changes to Editor, *Crime Laboratory Digest*, FSRTC, FBI Academy, Quantico, VA 22135.

146991

Guidelines for the Use of Mitochondrial DNA Sequencing in Forensic Science

Mark R. Wilson
Federal Bureau of Investigation
Washington, DC

Mitchell M. Holland
Armed Forces Institute of Pathology
Washington, DC

Mark Stoneking
Pennsylvania State University
University Park, PA

Joseph A. DiZinno
Federal Bureau of Investigation
Washington, DC

Bruce Budowle
Federal Bureau of Investigation
Quantico, VA

Human mitochondrial DNA (mtDNA) is an extrachromosomal, closed circular, organelle-specific genome consisting of approximately 16.5 kb. The mtDNA genome has been completely sequenced (Anderson *et al.* 1981). It consists of coding sequences for 2 ribosomal RNAs, 22 transfer RNAs, 13 proteins, and a noncoding region approximately 1,100 base pairs long, called the displacement loop (D-loop) or control region. Most of the sequence variation between individuals is found within the control region (Greenberg *et al.* 1983). The double-stranded mtDNA molecule is comprised of one purine-rich strand and one pyrimidine-rich strand, designated the heavy chain and light chain, respectively. Nucleotide positions in the mtDNA genome are numbered according to the convention of Anderson *et al.* (1981). The origin of replication of the heavy strand begins the numerical designation of each base pair, continuing around the circle for approximately 16,569 base pairs.

An important feature of mtDNA, which simplifies DNA sequencing, is its monoclonal nature. With the exception of certain disease conditions where tissue-specific deletions of large segments of the mitochondrial genome have been detected, for practical purposes all copies of an individual's mtDNA sequence are identical (Monnat and Loeb 1985; Monnat *et al.* 1985; Monnat and Reay 1986). Nuclear DNA, however, is diploid, consisting of a maternally inherited set of chromosomes and a paternally inherited set of chromosomes. In order to unambiguously sequence nuclear DNA (with the exception of the monoclonal Y chromosome), the two alleles must first be separated.

Due to its unique mode of inheritance, mtDNA does not undergo recombination. However, the low fidelity of mtDNA polymerase and the apparent lack of mtDNA repair mechanisms have led to a higher rate of mutation in the mitochondrial genome. Some regions of the mtDNA genome appear to be evolving at 5 to 10 times the rate of single copy nuclear genes (Brown *et al.* 1982; Cann *et al.* 1987). These areas of relatively high mutability are still well conserved. It is these regions which are of forensic interest, due to their ability to assist in the differentiation of individuals in the human population. It has been estimated that among Caucasians there is an average of 1 nucleotide difference every 100 bases (1%) in the most variable regions of mtDNA (Cann *et al.* 1987). This average is higher in individuals of African descent, being approximately 2.3% (Vigilant *et al.* 1991).

In the context of forensic science, the main advantage of typing mtDNA over nuclear DNA is its high copy number. While each set of nuclear chromosomes is present in only two copies per cell, mtDNA is present in hundreds or thousands of copies per cell. In cases where the amount of extracted DNA is very small or degraded, as in tissues such as bone, teeth, and hair, the probability of obtaining a DNA typing result from mtDNA is higher than that of polymorphic markers found in nuclear DNA.

In addition to its higher copy number and resultant increased sensitivity, mtDNA is maternally inherited (Case and Wallace 1981; Giles *et al.* 1980; Hutchinson *et al.* 1974). Barring mutation, the mtDNA sequence is identical for siblings and all their maternal relatives. This characteristic is helpful in forensic cases where known maternal relatives can provide reference samples for direct comparison to the questioned mtDNA type (Ginther *et al.* 1992; Holland *et al.* 1993; Stoneking *et al.* 1991).

An intriguing aspect of mtDNA which has recently been cited in the literature as a possibility is predicting the ethnic background of a DNA donor by his or her mtDNA sequence (Horai *et al.* 1993). It appears that specific mtDNA sequences are found in some ethnic groups but not in others. It may be possible to determine the ethnic background of an individual by his or her mtDNA sequence. Appropriately defined statistical estimates should always accompany such assessment of ethnicity.

Because of the potential for single nucleotide substitutions in mtDNA, a single base could represent the difference between a match and an exclusion in a forensic case. Therefore, careful guidelines for determining the means by which sequences are validated and reported must be established.

Mitochondrial DNA typing, using a variety of methods, has proven to be valuable as a means for human identification in many forensic cases (Higuchi *et al.* 1988; Hopgood *et al.* 1992; Orrego and King 1990; Stoneking *et al.* 1991). In some of these cases, mtDNA specific oligonucleotide hybridization assays have been employed (Stoneking *et al.* 1991); manual sequencing has been used in others (Ginther *et al.* 1992; Higuchi *et al.* 1988; Holland *et al.* 1993; Orrego and King 1990), and another group has reported research into fluorescence-based automated sequencing of mtDNA (Hopgood *et al.* 1992; Sullivan *et al.* 1992). In order to establish mtDNA typing as a method for large scale, routine forensic applications, it is proposed that some basic guidelines be adopted.

DNA Sequencing

Two distinct methods for sequencing DNA were published in the late 1970's (Maxam and Gilbert 1977; Sanger *et al.* 1977). The chain termination method, or Sanger method, is the more widely used. The Sanger method utilizes an oligonucleotide primer which is annealed to the DNA template to be sequenced. The primer is extended by a DNA polymerase in the presence of buffer, salts, deoxynucleotides (dNTPs), and dideoxynucleotides (ddNTPs). When a ddNTP is incorporated into the growing DNA strand, the chain can no longer be extended. Fragments of different lengths are thereby generated, each fragment's size depending on the exact site in which chain elongation is terminated. Generally one of the dNTPs is radiolabeled, and hence all sequencing fragments are internally labeled prior to separation in denaturing polyacrylamide gels. The result is a ladder of fragments, differing by a minimum of one nucleotide, which after autoradiography can be read as a sequence when all four termination reactions are run side-by-side. This type of sequencing, now called manual sequencing as opposed to automated sequencing, is comparatively more time consuming and labor intensive.

Polymerase chain reaction (PCR) amplified DNA templates can be sequenced either manually or through the use of an automated sequencer. Although manual sequencing is currently more widespread, automated sequencers provide high throughput and ease of data management through accompanying computer software. A number of different sequencing strategies have been employed for sequencing PCR-generated templates (Ellingboe and Gyllensten 1992; Gyllensten and Erlich 1988). Each method relies on the Sanger dideoxy terminator chemistry.

Fluorescent automated DNA sequencing was introduced in 1986 (Smith *et al.* 1986). In these sequencing reactions, fluorescently labeled oligonucleotide primers or dideoxy nucleotide terminators are incorporated into a DNA fragment which is then electrophoretically separated and detected by laser-induced fluorescence. The temporal, sequential passage of these fragments past the detector window gives rise to the sequence of the template. Since each terminator or primer can be uniquely labeled, all four termination reactions can be performed together (in the case of dye-labeled terminators) or performed separately and then pooled together (in the case of dye-labeled primers). In either case, the sequencing products are loaded on a sequencing gel in a single lane. Automated sequencing with fluorescent labels allows co-electrophoresis of each of the four termination reactions, thereby reducing the number of lanes which must be run per template from four to one.

Cycle sequencing is a technique which combines the dideoxy terminator chemistry with the sensitivity of PCR (Carothers *et al.* 1989; Murray 1989). The DNA templates are cyclicly denatured and annealed with sequencing primers which are extended in the presence of a thermally stable enzyme, such as *Taq* DNA polymerase. The reaction contains labeled primers, terminators, or a dNTP. Each fragment is uniquely labeled through the use of ddNTPs, as discussed previously. With each cycle the labeled fragments are denatured and removed from the template, which is again available for primer annealing, extension, and termination. Each template molecule is available for repeated use to generate a new sequencing product with each cycle. The sequencing product is thereby amplified, resulting in greater sensitivity with a limited amount of the DNA template. Because the DNA template is thermally denatured prior to primer annealing, double-stranded DNA can be used as a template in cycle sequencing reactions. This advantage obviates the need to generate single-stranded DNA from amplified PCR products prior to sequencing, thereby simplifying sample preparation. Cycle sequencing can be employed in both manual and automated formats.

Fluorescence-based automated sequencers have high throughput and permit rapid analysis of sequence information due to automated base calling. Through accompanying editing software, these instruments allow the user to rapidly compare sequences with one another, highlight the differences among multiple sequences, and compare individual sequences with a reference standard. For these reasons this approach is attractive to forensic science (Hopgood *et al.* 1992). However, caution must be exercised in using such instruments for reasons to be discussed later. One current manufacturer of automated DNA sequencers is Applied Biosystems, Inc. (Foster City, CA), which produces the automated DNA sequencer 373A. (Applied Biosystems, Inc. is a Division of Perkin Elmer, Norwalk, CT). The present discussion will be confined to the use of this instrument.

Two specific types of chemistries are employed with fluorescence-based sequencing: primer chemistry and terminator chemistry. With primer chemistry, the fluorescent label is attached to the 5' end of the sequencing primer. Four distinct labels are used, one for each dideoxy terminator reaction. For instance, a blue fluorescently tagged primer might be used with the cytosine dideoxy terminator, green might be used with an adenosine terminator, etc. Four separate terminator reactions are then performed on the DNA template to be sequenced, and the resultant products are pooled prior to electrophoresis. Terminator chemistry places the fluorescent label on the dideoxy terminator itself so that each of the four termination reactions can occur simultaneously in the same reaction vessel. Following a purification step to remove unincorporated dideoxy terminators, the products are directly loaded onto the sequencing gel contained within the instrument.

As electrophoretic separation of labeled fragments is progressing, real time detection, in the form of electronic signals, is sent to a computer to be analyzed. A laser scans the polyacrylamide gel plate once every 4 seconds. The electronic signals result from laser excitation and fluorescence of the separated fragments as they pass by a detector window. A photomultiplier tube within the instrument converts the optical signal into an electronic signal. A typical electrophoretic run lasts approximately 9 hours. Therefore, a tremendous amount of information, comprising about 12 megabytes, is generated from each sequencing run. This body of information is termed the gel file. It represents data points from across the gel collected at 4-second intervals over the duration of the run. The computer begins its analysis of the data from the information contained within the large gel file. It first assigns lanes to the gel file, which are detected as a series of positions within the gel where the fragments are detected. Once the lane positions are determined, the instrument analyzes the information within each lane, or file. Each lane is thereby tracked by the instrument, and information from each lane is placed into a file in the computer. Analysis of the DNA sequence is therefore dependent on the ability of the instrument to track lanes.

If the electric field through the gel is not uniform, the fragments will not migrate in a straight line. For example, if the gel is thicker on one side than the other, the resistance through the thick region will differ from the resistance through the thin segment. This difference could affect the ability of the instrument to accurately assign lanes and, in extreme cases, result in information from an adjacent lane being assigned to the wrong file.

In a typical sequencing run, the instrument accurately assigns lane positions. It also allows the user to manually override the lane positions by presenting an electronically generated gel window which shows the position of the fragments generated and where it has determined the lane position to be. The user is then free to realign the tracking position within a lane and to have the instrument reanalyze the data based on the new lane positions. It is obviously very important for the user to determine whether or not the tracking is indeed correct prior to any final analysis of the data. This can be done by examining the gel window. Appropriate positive control sequences also assist in data analysis. These are DNA templates of known sequence which can be analyzed together with unknown sequences. The known controls serve as sequencing controls and can also help determine the accuracy of the sequence analysis from run to run. An example of a sequence chromatogram from the Applied Biosystems, Inc. 373A DNA Sequencer is shown in Figure 1.

Data Formatting

The first entire human mitochondrial sequence was described by Anderson *et al.* (1981). This sequence is commonly referred to as the Anderson sequence, named after the first author of the 1981 publication, and it is often used as a standard sequence, by which other human sequences are compared. The Anderson sequence is listed as the light strand sequence, as opposed to the complementary heavy strand of the double-stranded native DNA. When a difference between an individual's sequence and the Anderson sequence is observed, the difference is sometimes referred to as a "polymorphism with respect to the Anderson sequence." For example, position number 16189 in the Anderson sequence lists the nucleotide thymidine or T. However, in some other individuals sequenced, that particular position contains a cytosine or C. If all other nucleotide positions were identical to the Anderson sequence, this individual's mtDNA type would be referred to as 16189-C, and it could be assumed that all other positions in the region sequenced other than 16189 are identical to the Anderson sequence.

Defined starting and stopping positions for mtDNA sequencing should also be a part of the mtDNA type. For example, in the case of the 16189-C individual, without defining the scope of the sequence considered, it might be assumed that the entire 16.5 kb of sequence exactly matched the Anderson sequence with one exception at position 16189. However, the vast majority of polymorphisms are found within two specific segments of the control region (Greenberg *et al.* 1983). These segments are termed hypervariable region 1 (HV1) and hypervariable region 2 (HV2). HV1 spans the approximate region of 16,024 to 16,365, and HV 2 is located on the other side of the origin of replication, approximately encompassing positions 73 to 340. The size of each of these regions makes them amenable for amplification by PCR and hence are useful for forensic purposes. Therefore, a possible way to list the mtDNA type of the individual in the earlier example could be to define the starting and ending points of the segment sequenced and then to list the polymorphisms with respect to the Anderson sequence in both HV1 and HV2 as follows:

HV1	HV2
16024-16365	73-340
16189-C	73 G
	146 C
	152 C
	199 C
	263 G
	315.1 C

As seen in this example, reporting a mtDNA type as the Anderson sequence (or some other standard sequence) as a reference and listing only deviations therefrom greatly simplifies processing and handling of mtDNA sequence data. In the future, it may be desirable to replace the Anderson sequence as a reference with a more common consensus sequence.

In addition to substitutions, insertions and deletions also are observed as polymorphisms in human mtDNA. It is proposed that insertions be indicated by placing a decimal point after the last base in the Anderson nomenclature and listing the insertion as the appropriate nucleotide thereafter. For example, length variations are observed to occur in HV2 close to a long stretch of cytosines interrupted by a single thymidine. The stretch of cytosines begins at position 303 and ends at position 315. The Anderson sequence contains an initial stretch of seven cytosines, followed by a single

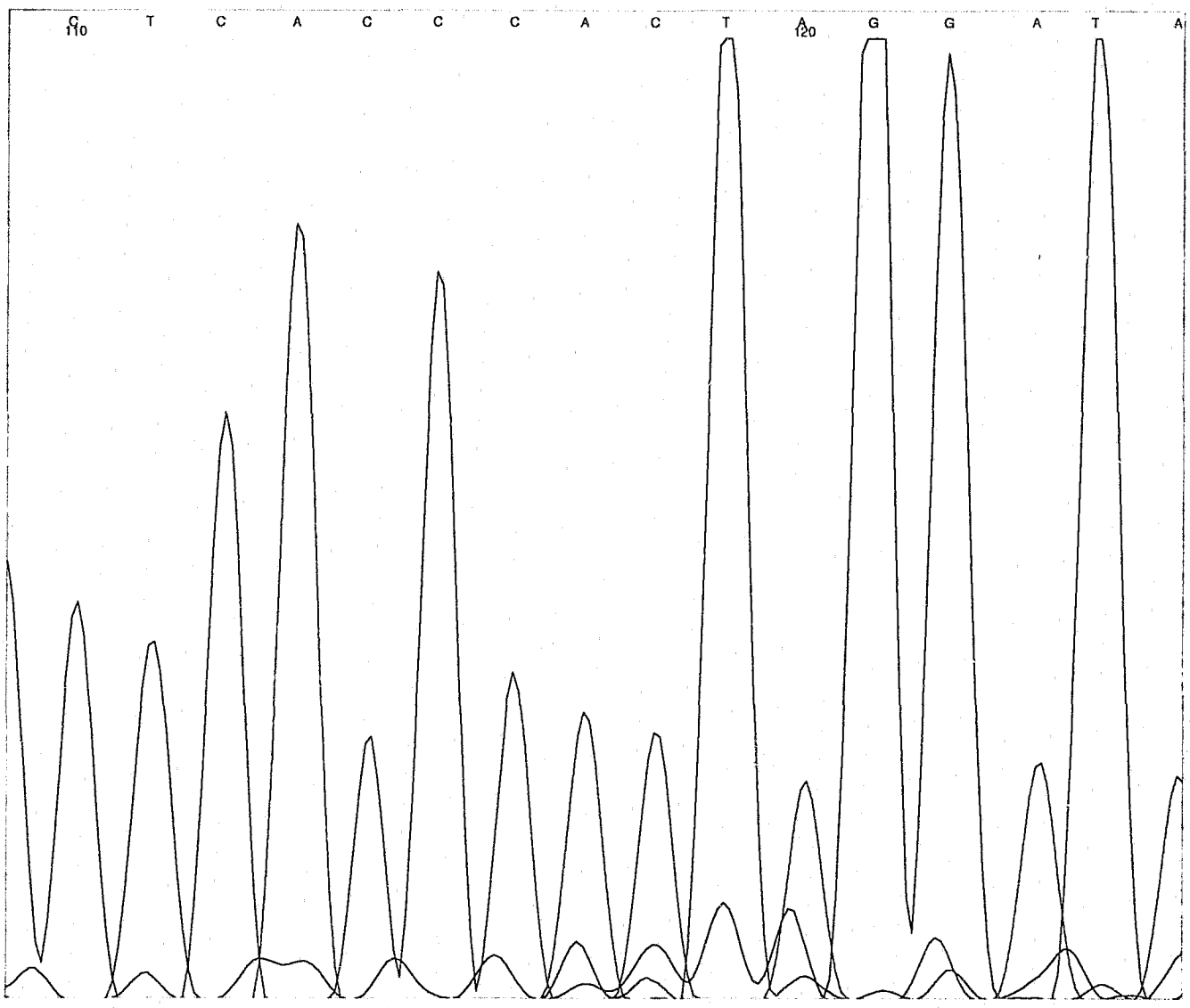


Figure 1. A limited portion of a mtDNA sequence obtained from HV1 of the control region of mtDNA through the cycle sequencing technique. The peaks represent the sequential passage by the detector window of mtDNA fragments that differ in length by one base. Fluorescence detection of the terminator dye on each fragment permits a determination of the final base incorporated into each fragment. The order of bases in this short sequence is shown at the top of the chromatogram. The result is the sequence of the template DNA, which in this case was a PCR product amplified from a human hair shaft extract.

thymidine and five additional cytosines. In some individuals, an extra cytosine is observed in the initial stretch of seven cytosines, resulting in eight cytosines prior to the sole thymidine. The insertion could, by convention, be recorded after the final cytosine in the initial run of seven and hence be designated as 309.1 C. Should the run of cytosines be longer than eight (for example, nine), prior to the thymidine, the type would be recorded as 309.1 C, 309.2 C. Deletions would simply be recorded as the number of the base or bases missing with respect to the Anderson nomenclature (for example, D315 or D315,316). Experience will determine whether this shorthand method of listing mtDNA types will prove to be a more effective manner of storing mtDNA sequence information.

Confirmation of DNA Sequences

Each sequence obtained should be confirmed, whether for databasing purposes or forensic casework. For example, if the light strand is sequenced and results in sequence A, the heavy strand can be sequenced and would result in sequence B. When the reverse complement of sequence B is compared with sequence A, the two should be identical. Alternatively, the same strand can be sequenced with the same or a different sequencing primer, independently confirming the initial sequence information from the same strand. Should any ambiguities arise between the original sequence and the confirmatory sequence, the DNA strand could be sequenced repeatedly until the ambiguity is

resolved (or the ambiguity could be designated as an N, meaning that the ambiguity is retained). Figure 2 shows a completed database sequence compiled using the Applied Biosystems, Inc. software package entitled SeqEd. Both strands of the template are sequenced, and the reverse complement of the heavy strand sequence is directly compared to the light strand sequence for confirmation.

The ultimate goal of forensic DNA sequencing should be to optimize each set of sequencing reactions so that no sequence ambiguities remain. However, when sequencing samples from forensic casework, complete, unambiguous sequencing results may not always be possible, even after confirmation. For example, PCR amplification of mtDNA extracted from liquid blood is generally robust and can result in large segments of amplified product. These large amplicons can be sequenced repeatedly using different sequencing primers or types of chemistries. As a result, one large amplicon could yield enough DNA template for repeated sequencing of different regions or the same region within the amplicon. Conversely, forensic samples such as bone and hair may not yield enough mtDNA to generate amplification of such large PCR products. However, these samples might yield short PCR fragments of approximately 200 to 300 base pairs. Sequencing these short amplicons may require alternate conditions, as the size of the DNA template is a variable to be considered in the process of optimizing a sequencing reaction.

Although it would be desirable to sequence an entire amplicon without any ambiguities, this may not be practicable. However, this is not necessary in order to acquire some information of forensic utility. One way that sequence ambiguities may be treated when compared to a mtDNA database will be discussed later.

Validation

A number of specific considerations should be addressed before mtDNA sequencing is used in forensic casework. They are as follows:

1. If database sharing among laboratories is to be accomplished, each sequencing strategy should yield identical results for the same DNA sequencing templates. For example, laboratories that sequence using manual techniques should share DNA templates with laboratories that utilize automated sequencers. Since automated sequencing can be done with dye terminators or dye primers, each of these methods should be compared and demonstrated to yield similar sequencing results on shared DNA templates within laboratories and between laboratories.
2. Maternal inheritance of mtDNA must be experimentally verified if it is to be used as a means of establishing identity in forensics. As maternal inheritance has been well documented, this verification should be minimal, encompassing a small number of families and showing that sequence information is identical from maternally related individuals.
3. A number of tissue types from the same individual should be used for DNA extraction and sequencing to show that an individual's mtDNA sequence is identical from tissue to tissue, regardless of the somatic origin of the extracted mtDNA. Studies with human hairs should show that different hairs from the same individual yield identical mtDNA sequences.
4. Environmental insult studies, such as exposure of samples to sunlight, soil, etc., should be performed to

```
BF 7026 HV2*
                        10         20         30         40         50         60         70         80
1 BF 7026-C GTGCACGCGA TAGCATTGCG AGACGCTGGA GCCGGAGCAC CCTATGTCGC AGTATCTGTC TTTGATTCCT GCCCCATCCC
2 BF 7026-D GTGCACGCGA TAGCATTGCG AGACGCTGGA GCCGGAGCAC CCTATGTCGC AGTATCTGTC TTTGATTCCT GCCCCATCCC
3           *--------- ---------- ---------- ---------- ---------- ---------- ---------- ---*------*
4 ANDERSON  ATGCACGCGA TAGCATTGCG AGACGCTGGA GCCGGAGCAC CCTATGTCGC AGTATCTGTC TTTGATTCCT GCCTCATCCT

                        90         100        110        120        130        140        150        160
1 BF 7026-C ATTATTTATC GCACCTACGT TCAATATTAC AGGCGAACAT ACTTACCAAA GTGTGTTAAT TAATTAATGC TTGTAGGACA
2 BF 7026-D ATTATTTATC GCACCTACGT TCAATATTAC AGGCGAACAT ACTTACCAAA GTGTGTTAAT TAATTAATGC TTGTAGGACA
3           ---------- ---------- ---------- ---------- ------*--- ---------- ---------- ----------
4 ANDERSON  ATTATTTATC GCACCTACGT TCAATATTAC AGGCGAACAT ACTTACTAAA GTGTGTTAAT TAATTAATGC TTGTAGGACA

                        170        180        190        200        210        220        230        240
1 BF 7026-C TAATAATAAC AATTGAATGT CTGCACAGCC GCTTTCCACA CAGACATCAT AACAAAAAAT TTCCACCAAA CCCCCCCTCC
2 BF 7026-D TAATAATAAC AATTGAATGT CTGCACAGCC GCTTTCCACA CAGACATCAT AACAAAAAAT TTCCACCAAA CCCCCCCTCC
3           ---------- ---------- ---------- *--------- ---------- ---------- ---------- ----------
4 ANDERSON  TAATAATAAC AATTGAATGT CTGCACAGCC ACTTTCCACA CAGACATCAT AACAAAAAAT TTCCACCAAA CCCCCCCTCC

                        250        260        270        280        290        300        310        320
1 BF 7026-C CCCCGCTTCT GGCCACAGCA CTTAAACAC
2 BF 7026-D CCCCGCTTCT GGCCACAGCA CTTAAACAC
3           ---*------ ---------- ---------
4 ANDERSON  CCC-GCTTCT GGCCACAGCA CTTAAACAC
```

Figure 2. Confirmatory sequence information can be generated by directly comparing the light strand (7026-C) of human mtDNA with the reverse complement of the heavy strand (7026-D) of the same region sequenced with a separate primer. Automatic reverse complementary alignment simplifies handling of sequence data and allows rapid confirmation of template sequence. Differences from the reference (Anderson) sequence are highlighted by an asterisk.

show that mtDNA sequence data is preserved and can be reliably typed. Studies on human hairs, for example, should include hairs exposed to dyes, bleaches, and other common hair treatments. As the chemistry of DNA is very well understood, it should be noted that if an extracted DNA sample enzymatically amplifies and sequences, if properly analyzed the sequence information correctly reflects that of the original extract.

5. With fluorescence-based automated DNA sequencers, a known positive control should be sequenced with each sequencing run. This yields information on the general quality of the sequencing reagents, the efficiency of unincorporated label removal, and the strength of the fluorescent signal. This lane thereby represents quality control for each sequencing run. This positive control should be of the same general physical form as the DNA sequenced from the unknown sample. This means that if double-stranded template DNA is used in a cycle sequencing format, a double-stranded sequencing control should be employed with each sequencing run.

Contamination

The most critical potential source of error in mtDNA sequencing is contamination. If more than one individual's DNA is extracted and amplified, the sequencing results will reflect this mixture. In extreme cases, the contaminating DNA can greatly exceed the DNA from the donor and thereby yield a false positive result. This fact would not be detected without the appropriate negative controls. Reagent blanks and negative controls should be employed with each extraction and amplification. Reagent blanks arise from an extraction procedure with no added biological material. Sterile water or buffer is carried through the extraction protocol, and the resultant extract is amplified as normal. This control tells the investigator if the extraction reagents are contaminated with human DNA. Negative controls are amplification-specific and test whether amplification reagents are contaminated. Sterile water or buffer is added to a PCR reaction in lieu of DNA and thermally cycled as the amplification protocol dictates.

The high sensitivity level of this method results in an increased possibility of contamination, which necessitates additional controls. If possible, based on the amount of tissue available for extraction, separate portions of the tissue should be separately extracted and amplified. This multiplicity gives the investigator an indication of the possibility of contamination and serves the additional purpose of independently confirming the sequence information from an independent extract of the same evidentiary material. In the case of a single evidentiary hair, however, duplicate extractions might not be possible. Future research should be directed toward determining the minimum amount of tissue necessary to obtain sufficient DNA for amplification. The sequencing results should be compared to the known mtDNA type of the donor (obtained from purified DNA from blood, for example) to show that the PCR-amplified amplicon from trace amounts of template exactly match that from more abundant sources of DNA template.

Population Frequency Calculations

Once a mtDNA type has been established, a determination of the frequency of this type in the general population could be presented in order for a jury to assess the weight of the evidence in a courtroom trial. General databases, which take into consideration the inherent genetic dissimilarities of the major races of man, should be compiled. These databases should include samples collected from the general population and generally should be divided into Caucasian, African, and Asian racial groups. An example of one form of such a database is shown in Figure 3 (Stoneking *et al.* 1991).

As the mitochondrial genome is inherited in its entirety without recombination, a complete mtDNA sequence is the equivalent of a single genetic locus. Therefore, the frequency of a mtDNA type might be multiplied by allele frequencies of polymorphic nuclear loci, but at this time mtDNA sequence information from within an individual cannot be multiplied with other mtDNA sequence information. For example, since HV1 and HV2 are genetically linked due to their mode of inheritance, it would not be appropriate to multiply the frequency of a mtDNA sequence from HV1 with that of HV2. HV1 and HV2 together constitute a single mtDNA type. Therefore, at this time the frequency of a mtDNA type within the population is calculated by the counting method, or the number of times a particular sequence has been observed in a population, divided by the number of sequences contained within the appropriate database.

In forensic casework, it may not always be possible to amplify and/or sequence the entire regions defined as HV1 and HV2. In some cases, the size of the amplified product could be significantly smaller. In these cases, to obtain a statistical estimate it is necessary to precisely define the starting and stopping points of the sequence to be queried with respect to the database. For example, if data on only 100 base pairs within HV1 from position 16100 to position 16200 are available, the database should be queried for a frequency only within these two points. The remaining information in the database should not be used for statistical purposes in this particular case.

As discussed previously, forensic sequence information might not be without ambiguity at certain positions. In such cases, all four nucleotide frequencies in the reference population might be considered at each position of ambiguity. For example, consider the mtDNA sequence from positions 80 to 100 (HV2). The sequence from Anderson *et al.* (1981) follows: CGA TAG CAT TGC GAG ACG CTG.

Suppose a forensic sample is amplified and sequenced from positions 80 to 100. The forensic sample is sequenced in duplicate, but at position 86 the base cannot be determined. The forensic sample is also found to contain a polymorphism with respect to the Anderson sequence at position number 98. Instead of a cytosine at this position, a thymidine is found. One way to search the mtDNA database for such a sample might be to query for the frequency of the following sequence beginning at position 80 and ending at position 100:

80 100
CGA TAG NAT TGC GAG ACG TTG

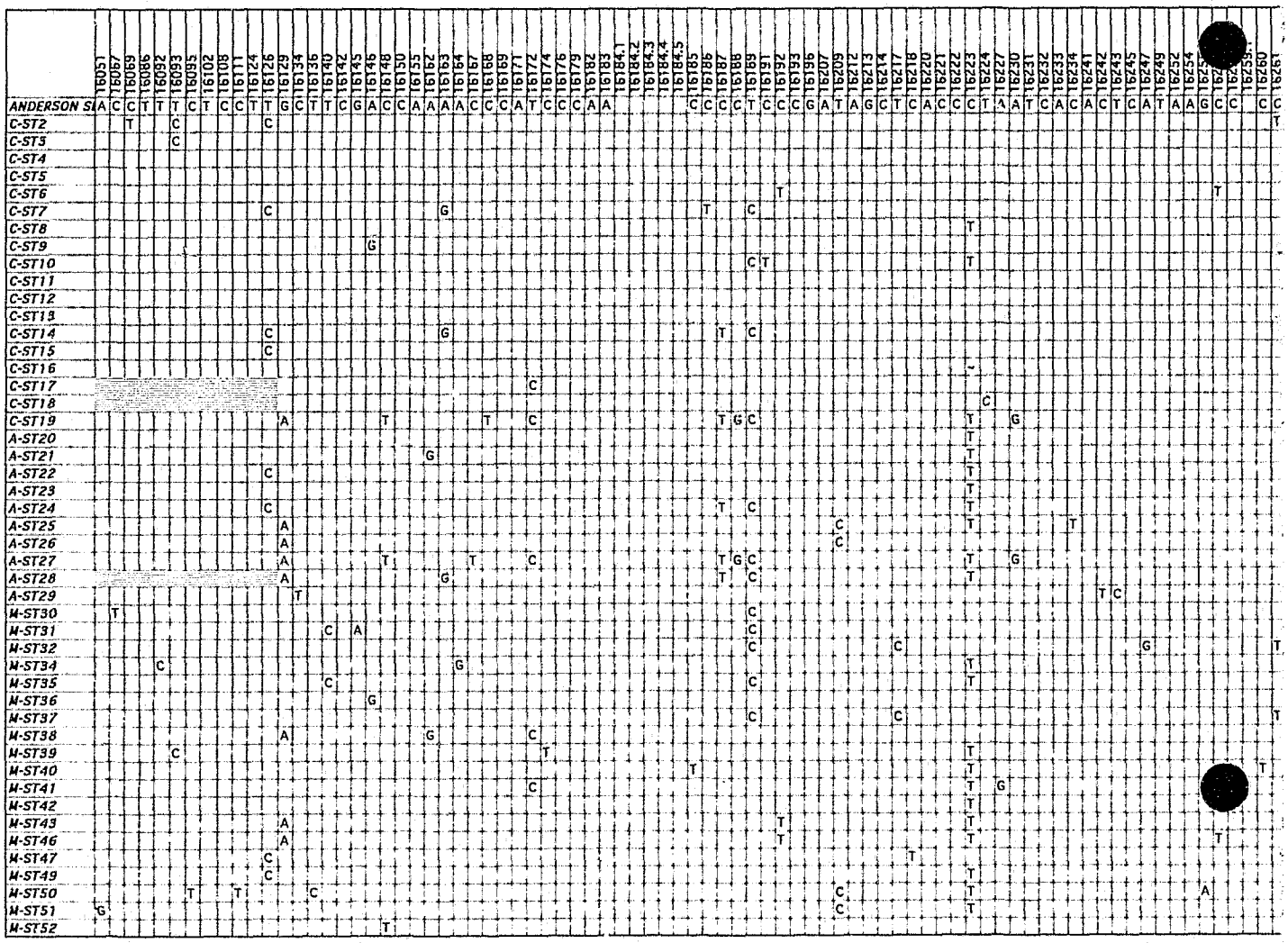
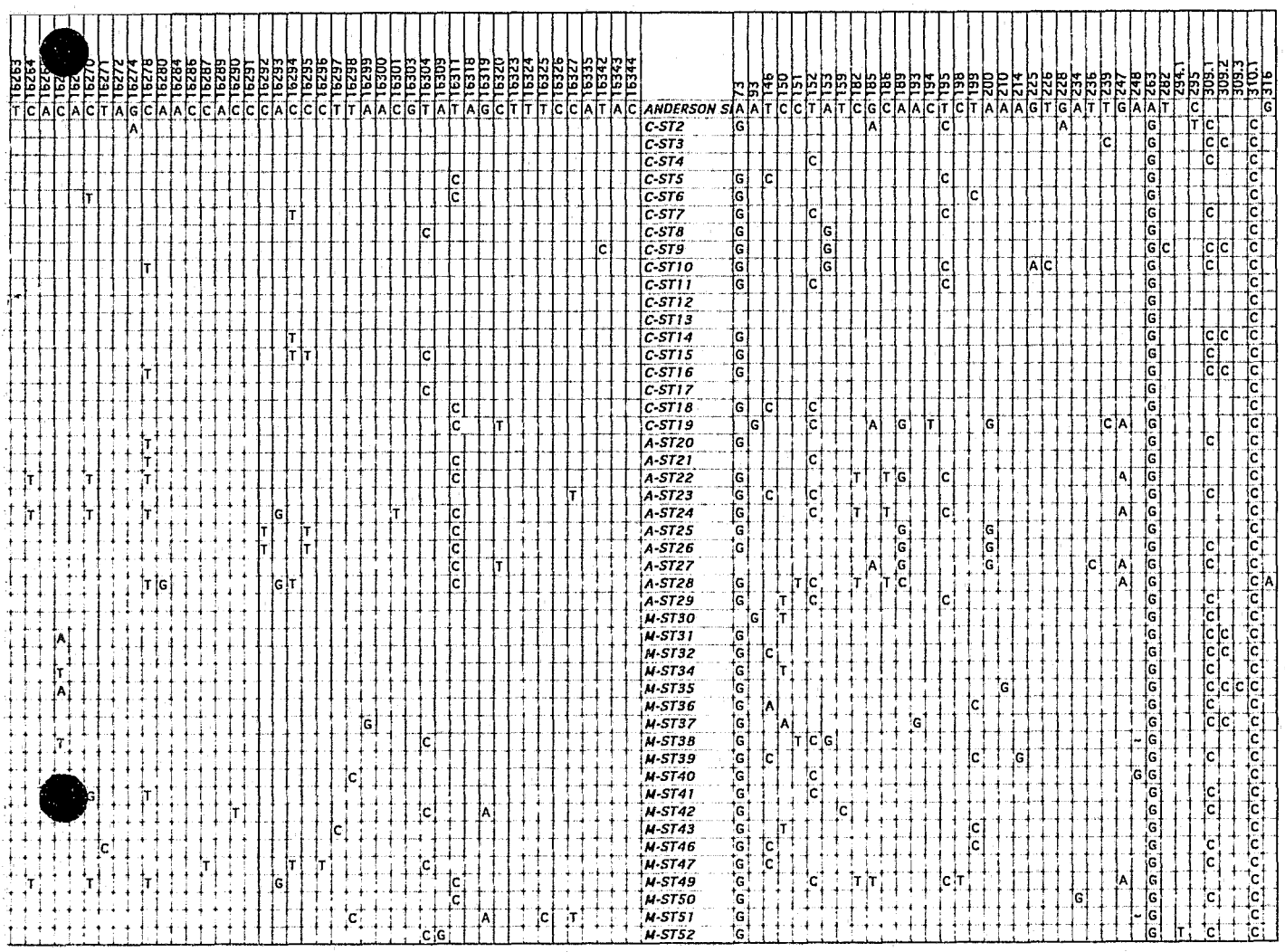


Figure 3. Upon completion of sequence confirmation, multiple sequences can be aligned and presented together. This data is from Stoneking et al. (1991). The standard sequence (Anderson) is shown on the first line. Deviations from this sequence are shown as the corresponding letter of the nucleotide which differs from the reference. Where the boxes are empty, the sequence of the sample matches that of the reference. Where a polymorphism with respect to the reference (Anderson) sequence exists, the differing nucleotide is placed



in the appropriate box. Transversions are printed in boldface. Horizontal lines denote sequence information which was not determined. Deletions are represented by a dash. Notice that in some positions, such as nucleotide positions 263 and 310.1 in HV2, the reference nucleotide is rare in the population, and a better way to report the information might be to establish a different reference sequence, one which represents the most common base at each position.

The N at position 86 tells the computer to allow any of the four possibilities at this position. The T at position 98, in addition to all the other positions, is fixed. Then, in effect the computer sums the frequencies of the following sequences:

CGA TAG AAT TGC GAG ACG TTG
CGA TAG CAT TGC GAG ACG TTG
CGA TAG GAT TGC GAG ACG TTG
CGA TAG TAT TGC GAG ACG TTG

Discussion

Mitochondrial DNA sequencing is a viable technique for human identification testing. Its advantage over nuclear DNA typing methods is the high copy number of mtDNA over nuclear DNA and hence the added sensitivity in cases where one is confronted with a limited amount of DNA or the DNA is significantly degraded. The maternal inheritance of mtDNA offers the forensic scientist reference samples from maternally related individuals which are, barring mutation(s), identical. Nuclear markers from related individuals share common alleles but are not genotypically identical.

As with all forensic identification techniques, issues related to reliability and reproducibility must be addressed. Automated sequencers offer high throughput and ease of data handling but also introduce issues related specifically to this technology. Replication of sequencing runs, which independently confirm nucleotide designations, is a good way of overcoming sequence ambiguities. As many different methods exist to sequence DNA, templates should be shared among laboratories, and the results should be independently assessed for reliability.

Validation studies must be conducted on any new forensic technique. As the chemistry of DNA is very well understood, these studies should be confined to examining the direct influence of common environmental agents (*e.g.*, dyes and bleaches) on human hairs.

Statistical interpretation of mtDNA sequence information is straightforward. A calculation of the frequency of a sequence is simply the number of times the particular sequence has been observed within a specified region, divided by the number of samples in the appropriate database. Where sequencing ambiguities arise, these positions can be considered nonfixed, and all possibilities can be included in the statistical calculation. However, meaningful frequency estimations require that reference databases containing DNA sequence information be free from ambiguities within the region of comparison.

We propose that some basic guidelines be adopted in order to facilitate the widespread use of mtDNA sequencing in forensic science. These guidelines are as follows:

1. A common nomenclature should be adopted which simplifies recording sequence polymorphisms by listing only deviations from a reference sequence and assumes all other positions are identical to the reference.
2. Extracts should be amplified in duplicate whenever feasible.
3. Amplicons should be sequenced in duplicate, either by sequencing the complementary strand of the duplex DNA molecule or by additional sequencing of the same strand.
4. A common nomenclature for primer designation should be adopted. Presently, mtDNA primers are listed by the strand (H for heavy and L for light) followed by the numeric designation from Anderson *et al.* (1981) of the 3' ending base of the primer.
5. Laboratories should share template DNAs for interlaboratory comparison of sequence information. This will facilitate validation of numerous sequencing techniques, ensuring that quality control can be maintained regardless of a laboratory's methodology. Independent, nonforensic laboratories should be included in this effort to implement consensus quality control measures.
6. A human mtDNA database from HV1 and HV2 of the D-loop should be compiled from existing published sequences and ongoing research. When feasible, DNA samples from contributors of published mtDNA control region sequences found in the literature should be obtained and sequenced by forensic laboratories to ensure interlaboratory quality control and fidelity of the forensic mtDNA database.

It should be stressed that mtDNA sequencing is an analytical procedure which the forensic scientist must understand in its proper context. In routine forensic cases where nuclear DNA is extracted from a sample, polymorphic nuclear markers such as DQα, short tandem repeat loci, etc., can and should be utilized. These markers should take precedence over mtDNA analysis. Not only are such markers currently more statistically informative than mtDNA, they also are not as susceptible to contamination as the analysis of mtDNA. However, mtDNA can and should be the method routinely used with certain classes of biological evidence which by nature have very little DNA, such as telogen hairs, hair shafts, bones, and teeth.

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.

Brown, W. M., Prager, E. M., Wang, A., and Wilson, A. C. Mitochondrial DNA sequences of primates: Tempo and mode of evolution, *Journal of Molecular Evolution* (1982) 18:225-239.

Cann, R. L., Stoneking, M., and Wilson, A. C. Mitochondrial DNA and human evolution, *Nature* (1987) 325:31-36.

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.

Ellingboe, J. and Gyllensten, B., eds. *The PCR Technique: DNA Sequencing*. Eaton Publishers, Natick, MA, 1992.

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 of the United States of America* (1980) 77:6715-6719.

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.

Gyllensten, U. B. and Erlich, H. A. Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus, *Proceedings of the National Academy of Sciences of the United States of America* (1988) 85:7652-7656.

Higuchi, R., von Beroldingen, C. H., Sensabaugh, G. F., and Erlich, H. A. DNA typing from single hairs, *Nature* (1988) 332:543-546.

Holland, M. M., Fisher, D. L., Mitchell, L. G., Rodriguez, 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.

Hopgood, R., Sullivan, K. M., and Gill, P. Strategies for automated sequencing of human mitochondrial DNA directly from PCR products, *Biotechniques* (1992) 13:82-92.

Horai, S., Kondo, R., Nakagawa-Hattori, Y., Hayashi, S., Sonoda, S., and Tajima, K. Peopling of the Americas, founded by four major lineages of mitochondrial DNA, *Molecular Biology and Evolution* (1993) 10:23-47.

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

Maxam, A. M. and Gilbert, W. New method for sequencing DNA, *Proceedings of the National Academy of Sciences of the United States of America* (1977) 74:560-564.

Monnat, R. J. and Loeb, L. A. Nucleotide sequence preservation of human mitochondrial DNA, *Proceedings of the National Academy of Sciences of the United States of America* (1985) 82:2895-2899.

Monnat, R. J., Maxwell, C. L., and Loeb, L. A. Nucleotide sequence preservation of human leukemic mitochondrial DNA, *Cancer Research* (1985) 45:1809-1814.

Monnat, R. J. and Reay, D. T. Nucleotide sequence identity of mitochondrial DNA from different human tissues, *Gene* (1986) 43:205-211.

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

Orrego, C. and King, M. C. Determination of familial relationships. In: *PCR Protocols: A Guide to Methods and Applications*, ed. M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White, 416-426. Academic Press, San Diego, CA, 1990.

Sanger, F., Nicklen, S., and Coulson, A. R. DNA sequencing with chain-terminating inhibitors, *Proceedings of the National Academy of Sciences of the United States of America* (1977) 74:5463-5468.

Smith, L. M., Sanders, J. Z., Kaiser, R. J., Highes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. H., and Hood, L. E. Fluorescence detection in automated DNA sequence analysis, *Nature* (1986) 321:674-679.

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.

Sullivan, K. M., Hopgood, R., and Gill, P. Identification of human remains by amplification and automated sequencing of mitochondrial DNA, *International Journal of Legal Medicine* (1992) 105:83-86.

Vigilant, L. M., Stoneking, M., Harpending, H., Hawkes, K., and Wilson, A. C. African populations and the evolution of human mitochondrial DNA, *Science* (1991) 253:1503-1507.