1996-IJ-CX-0028

Development of a Rapid, Immobilized Probe Assay for the Detection of Mitochondrial DNA Variation in the HVI and HVII Regions

Cassandra D. Calloway, MS and Henry Erlich, PhD

Abstract

Since the human mitochondrial genome was first sequenced in 1981 by Anderson et al., the sequence diversity of the mitochondrial genome has been well characterized and mitochondrial DNA (mtDNA) has proven to be a useful target for the analysis of forensic materials. With the support of the NIJ, we have developed a rapid method for analysis of sequence variation in the HVI and HVII regions of the human mitochondrial genome that utilizes the established technologies of PCR amplification and immobilized probe hybridization. For the initial version, a subset of the 23 Sequence-Specific Oligonucleotide (SSO) probes developed for a mtDNA typing system that used the standard dot blot format (Stoneking et al. 1991) was adapted to the more convenient and well accepted immobilized probe method ('reverse dot blot' or linear array). This first linear array probe panel for mitochondrial DNA sequence analysis was comprised of 17 SSO probes targeting the HVII region of the control region. To increase the flexibility and discrimination power of this assay, we incorporated a primer pair for the HVI region into our PCR amplification reaction and added 18 additional probes to detect sequence variation in four regions of HVI and two additional heteroplasmy hotspots (16093 and 189). The current commercially available HVI/HVII linear array probe panel consists of two primer pairs for co-amplification of HVI and HVII PCR products and 33 probes immobilized in 31 lines for detection of sequence variation at 18 positions spanning both hypervariable regions. Developmental validation of the LINEAR ARRAY mtDNA HVI/HVII Region-Sequence Typing Kit has been completed and internal validation within forensic DNA testing laboratories is on-going. Throughout the course of development, this assay was extensively evaluated within our laboratory as well as nearly 20 laboratories who participated in beta site studies or collaborative projects. Population studies were conducted to evaluate the sequence diversity and mtDNA type

frequencies within and between several populations. Further, heteroplasmy for several tissue types, including hair, muscle, and blood was characterized. Mixed and degraded samples as well as simulated and non-probative casework samples were tested. In addition, the assay has been used successfully in Sweden for analysis of more than 300 forensic casework samples (Allen et al. 1998; Divne et al. manuscript in preparation) and for the identification of mass grave remains in Croatia (Gabriel et al. 2003). From these analyses, it was concluded that the HVI/HVII linear array probe panel is an effective tool for analyzing forensic casework samples and for reducing the number of samples that require sequencing.

Executive Summary

Introduction

The human mitochondrial genome was first sequenced in 1981 by Anderson et al. and subsequently the sequence diversity of the mitochondrial genome has been well characterized. Mitochondrial DNA (mtDNA) has several properties that make it particularly useful for human identification, including high copy number per cell, maternal inheritance, and high sequence variability. Today, an increased number of cases are submitted for mtDNA analysis and the number of laboratories performing mtDNA analysis continues to increase (Wilson et al. 1995; Allen et al. 1998; Carracedo et al. 2000; Melton and Nelson 2001). To date, the most widely used method for mtDNA analysis is DNA sequencing. The sequencing strategy commonly used requires two to four amplification reactions and four to eight sequencing reactions, depending on the quality of DNA.

Our goal was to develop a sensitive, robust amplification and typing system to detect most of the diversity revealed by the HVI and HVII regions of the control region of the human mitochondrial genome, but at a fraction of the time, effort, and cost of actually sequencing these regions. We successfully developed and made commercially available the LINEAR ARRAY mtDNA HVI/II Region-Sequence Typing Kit (Roche Applied Science, Indianapolis, IN), a product which uses the established technologies of PCR amplification and immobilized probe hybridization (‘reverse dot blot’ or linear array). We chose to develop a mtDNA typing system that uses a set of SSO probes immobilized in lines on a strip of nylon membrane, since this method of typing is rapid, results are easy to interpret, and linear array probe panels are easily manufactured. Also, this method for typing mtDNA sequence variation can be performed in a single day using established DNA typing technologies and it can be used to quickly generate large population databases.

Another advantage to the linear array technology is that it requires only a thermal cycler and a rotating water bath and can be implemented by any laboratory since it is simple to adopt and provides a low-budget solution to mtDNA typing compared to DNA sequencing. For high-throughput laboratories, automated instrumentation can be used to improve throughput. In addition, significant advances in automating the 'reading' of the linear arrays have been made.

Over the course of the granting period, we developed several mtDNA typing systems including a HVII linear array with a panel of 17 probes (Reynolds et. al. 2000; Calloway et. al. 2000), a HVI/HVII linear array with a panel of 27 probes (Gabriel et. al. 2001; Gabriel et. al. 2003), and the commercially available LINEAR ARRAY mtDNA HVI/II Region-Sequence Typing Kit (Roche Applied Science, Indianapolis, IN) with a panel of 33 probes (Kline et. al. 2005; Divne et al. 2005). We describe below the various versions of the assay leading up to the final, commercially available system.

The first 'reverse line blot' assay (linear array) for mitochondrial DNA sequence analysis was comprised of 17 SSO probes targeting the HVII region of the control region. The standard dot blot system developed with Dr. Mark Stoneking (1991) was used as a starting point for the reverse hybridization assays, although a majority of the probes had to be redesigned to work optimally together under the conditions of the immobilized probe method. The HVII region was chosen over the HVI region as a target for the first assay since it had the greater potential for a higher discrimination power versus the HVI region. To decrease the number of blanks observed with the Stoneking assay and to further increase the power of discrimination for the HVII region, several additional probes were incorporated into the HVII probe panel including four probes

detecting sequence variants in the B region and two probes detecting sequence variants in the C region. Data showed that the additional probes significantly decreased the observed number of blanks, increasing the power of discrimination compared to the original set of HVII probes used with the Stoneking assay (Reynolds et al. 2000). However, the discrimination power of the HVII probe panel was less than the 23 SSO probes targeting both HVI and HVII regions.

To increase the power of this assay, we incorporated a primer pair for the HVI region into our PCR amplification reaction and initially added probes to detect sequence variation in two regions of HVI (IA and IC) and at two heteroplasmy hotspots (16093 and 189). Probes detecting 189A/195C and 189A/195T are striped in a single line as well as probes detecting 189G/195C and 189G/195T to conserve space on the array and since the IIC region probes detect variation at 195. The "C-stretch" probe that detected length variation in the poly-C tract and the intensity control probe from the HVII array were removed. The C-stretch probe was removed because of the high frequency of length heteroplasmy in this region as well as the potential for cross-hybridization if HVI length heteroplasmy is present in the sample. This intermediary version of the HVI/HVII linear array was developed and used to establish a population database of 105 Croatians (Gabriel et al. 2001) and to identify victims of mass graves in Croatia (Gabriel et al. 2003).

To further increase the discrimination power of the assay, probes targeting two more regions of the HVI region were added, ID as defined by Stoneking et al (1991) and IE which was identified through sequence analysis. Probes targeting 16270 and 16278 (region IE) were chosen over probes targeting region IB, because the addition of the IE probes increased the discrimination power among Caucasians (for which the power of discrimination was among the lowest of the population groups tested). This version consists of 33 probes immobilized in 31 lines for

detection of sequence variation at 18 positions spanning both hypervariable regions and is commercially available through Roche Applied Science as the LINEAR ARRAY mtDNA HVI/HVII Region-Sequence Typing Kit. These sites and probe designations are described in

Table 1.

HVI Probe Designations	Sequence Variation Detected												
			16093										
16093 1	A	T	**T**	T	C								
16093 2	·	·	**C**	·	·								
			16126			**16129**							
IA1	T	G	**T**	A	C	**G**	G	T					
IA2	·	·	**C**	·	·	·	·	·					
IA3	·	·	·	·	·	**A**	·	·					
			16304					**16309**		**16311**			
IC1	A	G	**T**	A	C	A	T	**A**	G	**T**	A	C	
IC2	·	·	**C**	·	·	·	·	·	·	·	·	·	
IC3	·	·	·	·	·	·	·	·	·	**C**	·	·	
IC4	·	·	·	·	·	·	·	**G**	·	·	·	·	
ICw2/w3	·	·	**C**	·	·	·	·	·	·	**C**	·	·	
			16362										
ID1	C	G	**T**	C	C								
ID2	·	·	**C**	·	·								
			16270								**16278**		
IE1	C	A	**C**	T	A	G	G	A	T	A	**C**	C	A
IE2	·	·	·	·	·	·	·	·	·	·	**T**	·	·
IE3	·	·	**T**	·	·	·	·	·	·	·	·	·	·

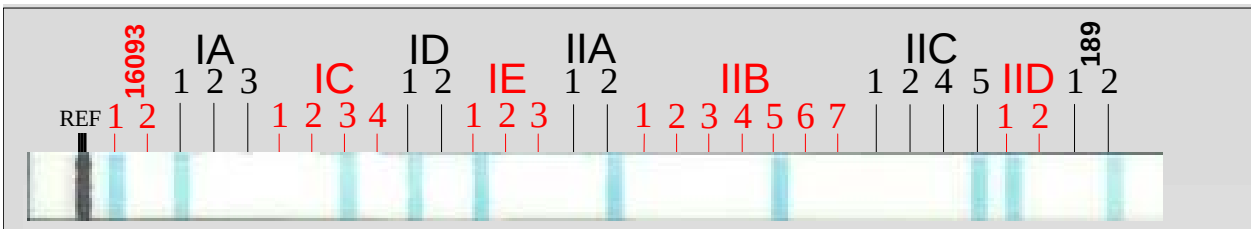
Table 1 (continued)

HVII Probe Designations	Sequence Variation Detected										
			73								
IIA 1	G	T	**A**	G	T						
IIA 2	·	·	**G**	·	·						
			146				**150**		**152**		
IIB1	C	C	**T**	C	A	T	**C**	C	**T**	A	T
IIB2	·	·	**C**	·	·	·	·	·	·	·	·
IIB3	·	·	·	·	·	·	·	·	**C**	·	·
IIB4	·	·	**C**	·	·	·	·	·	**C**	·	·
IIB5	·	·	·	·	·	·	**T**	·	·	·	·
IIB6	·	·	·	·	·	·	**T**	·	**C**	·	·

IIB7	·	·	**C**	·	·	·	**T**	·	**C**	·	·								
			189						**195**			**198**		**200**					
IIC1	G	A	**A**	C	A	T	A	C	**T**	T	A	**C**	T	**A**	A	A			
IIC2	·	·	·	·	·	·	·	·	**C**	·	·	·	·	·	·	·			
IIC4	·	·	·	·	·	·	·	·	**C**	·	·	**T**	·	·	·	·			
IIC5	·	·	**G**	·	·	·	·	·	·	·	·	·	·	**G**	·	·			
			247																
IID1	T	T	**G**	A	A														
IID2	·	·	**A**	·	·														
			189						**195**										
189 1	G	A	**A**	C	A	T	A	C	T	T	A								
	·	·	.	·	·	·	·	·	C	·	·								
189 2	·	·	**G**	·	·	·	·	·	.	·	·								
	·	·	**G**	·	·	·	·	·	C	·	·								

The mtDNA linear array typing results for one individual (with mitotype 1131125512) are shown below in Figure 2.

Figure 2.



The mtDNA HVI/II linear array can be used as a screening tool to eliminate suspects quickly and to identify samples requiring further genetic analysis (i.e. sequencing). For example, suspects and specimens (e.g., multiple hairs) associated with a specific case can be rapidly screened with immobilized SSO probes. Samples from suspects who cannot be excluded and specimens that match the reference sample by the immobilized probe linear array could then be sequenced if desired without having to generate additional PCR product. Unlike standard amplification procedures for mtDNA sequencing analysis, the HVI/II regions are co-amplified, thereby

reducing the amount of DNA material consumed during amplification by 50 to 75% as well as PCR set-up time. In addition, less than 50% (~20 of the 50 μL) of the PCR product generated with the assay is consumed during the quantification and hybridization steps, leaving the remainder of the PCR product (enough for forward and reverse sequencing of both HVI and HVII regions) for subsequent sequence analysis if needed. High quality DNA sequence has been consistently obtained using PCR product generated from the duplex reaction (Date Chong et al. 2005). The very low to no background noise in the sequencing chromatogram is likely a result of the high specificity of the optimized duplex amplification reaction. This PCR system has been optimized to ensure both products amplify with approximately equal efficiency, primer dimer formation is minimized, and non-specific products are rarely generated. A higher amount of TaqGold DNA polymerase is used in the amplification to increase sensitivity and reduce the effect of potential PCR inhibitors.

Probes were optimized to minimize cross-hybridization, increase sensitivity and give approximately equal signal intensity within a hypervariable region. Typing conditions were optimized for both a rotating water bath for manual typing and the Tecan Profiblot for automated typing. To improve throughput of data analysis, software and hardware for scanning the developed “strips” and inferring a genotype are also available for the commercially available CF and HLA tests. Some modifications of the existing scanning and interpretation software will be required for the mtDNA typing system and are underway.

Developmental Validation and Collaborative Studies

Before a new DNA typing system can be used by the forensic community, it must undergo developmental validation. This involves optimizing the test and evaluating its performance to establish its capabilities. Several improvements to the amplification and typing assay were made during optimization, including shortening the HVII reverse primer to further minimize primer dimmer formation and using a single buffer for both hybridization and wash steps during typing. Once the systems were optimized and finalized, experiments designed to identify acceptable ranges in reaction parameters ('window studies') were completed and equivalent results were observed for the specified range for each parameter tested necessary for manufacturing allowances.

To complete the developmental validation, a series of studies were conducted to evaluate the performance of the assay as suggested by the Technical Working Group on DNA Analysis Methods (TWGDAM) committee and DNA Advisory Board (DAB). Results from a subset of these studies are summarized below.

Species specificity of the mtDNA LINEAR ARRAY Amplification and Typing system was examined using nonhuman DNA extracts. Three concentrations (500, 50, and 5 pg) of DNA samples from 16 species were amplified and typed: primates (gorilla, chimpanzee, orangutan, marmoset) and non-primates (mouse, rat, dog, cat, fish, cow, sheep, chicken, turkey, pig, e-coli, candida). Only products for higher order primates hybridized (gorilla, chimpanzee, and orangutan) and only to a subset of probes in the panel.

Sensitivity studies were conducted and optimal intensity of probe signals on the linear arrays was observed when roughly 1.0 pg or more of total genomic DNA was amplified for 34 cycles. Typing of less than 0.5 pg of DNA sample required 38 cycles of amplification to obtain enough amplicon for optimal signal with the array. Stochastic effects were observed for some samples that were amplified for 38 cycles with 0.05 pg of template DNA.

A series of mixture studies have been conducted and the minor component was consistently detectable down to 10:1 ratio. In one such study conducted at the California Department of Justice, two DNA samples that differ at six positions were mixed in defined ratios and typed. The results indicated that the minor component was detectable down to a 10:1 ratio (the lowest level tested in this study). Additional mixtures studies have been conducted and in some cases the minor component was detectable down to a 20:1 ratio.

Additionally, a series of population studies was conducted to determine the distribution of the various mitotypes and to calculate the level of discrimination power for the linear array probe panel for a number of population groups. Using a panel of 17 SSO probes immobilized on nylon membrane strips targeting the HVII region, blood samples from 689 individuals from four population groups were typed (Reynolds et al. 2000). The genetic diversity value for each population was calculated from the frequency data collected for the HVII probe panel and range from ~0.95 to 0.98, highest among African Americans. 674 available samples typed previously with the HVII array were typed using the final panel of 33 probes targeting 18 sites within the HVI and HVII regions (Calloway et al. manuscript in preparation). The genetic diversity values ranged from ~9.5-9.9 with the final 33 probe panel (see Table 1). The highest discrimination

power was observed for the African Americans and the lowest discrimination power was observed for the Hispanic and Caucasians groups. A large number of individuals having a few common types account for the lower discrimination power in these two populations. This set of samples was also sequenced and sequence analysis had a limited ability to further distinguish individuals among some common mitotypes. Thus, the greatest limitation of both HVI/II sequence *and* linear array analyses lies with the small number of common types. Variation outside the HV regions will need to be targeted to further discriminate individuals having these common types. Although several common mitotypes were observed, most of the mitotypes occurred only once within the population group (see # unique mitotypes in Table 1).

Additionally, blood samples randomly collected from 100 Caucasians and 100 African Americans from individuals who resided in Georgia, were typed with the 33 probe panel. The HVI and HVII regions of each of the 200 samples were sequenced as well. The LINEAR ARRAY™ mtDNA HVI/HVII Region-Sequencing Typing Kit (Roche Applied Science, Indianapolis, IN) was used also to type 666 individuals from U.S. Caucasian, African American, and Hispanic groups. Processing of the LINEAR ARRAY probe panels was automated on a Tecan ProfiBlot workstation (Kline et al. 2005).

Throughout the course of development, this assay was extensively evaluated within our laboratory as well as nearly 20 laboratories who participated in beta studies or collaborative projects. These studies included extensive analysis of case type samples. Samples exposed to a range of environmental conditions (e.g. snow, heat, rain) and deposited on a wide variety of substrates (e.g. carpet, clothing, asphalt, dirt) were typed at the San Bernardino Sheriff's Office

(Anjaria 2003). To investigate the value of the linear arrays as a screening tool, reference hairs and blood samples from victims and suspects were obtained from rape kits collected for sexual assault cases, assembled into eight mock cases and analyzed by two laboratories (Anjaria 2003). In addition, samples from multiple adjudicated cases were typed by the Georgia Bureau of Investigation using the SSO linear array and sequence analysis (Williams et al. manuscript in preparation). The mtDNA array system has been used successfully in Sweden for analysis of more than 300 forensic casework samples (Allen et al. 1998; Divne et al. 2005). From these analyses, it was concluded that the linear array assay is an effective tool for analyzing the origin of hairs shed at a crime scene and reducing the number of samples requiring sequencing.

Sequence variation detection for 105 randomly selected Croatian individuals with an array of 27 immobilized SSO probes that targets six regions within HVI and HVII and two additional sites (189 and 16093) was used to establish a reference database for mtDNA analysis of human remains recovered from mass graves (Gabriel et al. 2001). The system also has been used successfully to identify remains from a mass grave site in Croatia (Gabriel, Calloway et al. 2001; Gabriel et al. 2003). We analyzed the mtDNA control region of 18 skeletal elements recovered from a mass grave in Croatia that could not be typed using nuclear markers. One set of remains was positively identified using a linear array probe panel and confirmed by sequence analysis.

Heteroplasmy was also extensively studied in our laboratory. In the first study, 459 hairs from 23 people were analyzed using the HVII linear array. Point heteroplasmy was observed in one or more hairs in 22% of individuals. In contrast, only 0.8% of the 689 blood samples from individuals typed using the HVII array to generate the population data base were heteroplasmic

(point). To investigate the frequency of heteroplasmy across various tissues and age groups, the HVII array was used to screen for heteroplasmy from heart, brain, muscle, and blood samples from 43 cadavers (Calloway 1998; Calloway et al. 2000). Results showed (a) higher levels in muscle tissue than in the other tissues, blood showing the least, and (b) higher levels in tissues from older individuals (Calloway 1998; Calloway et al. 2000). Subsequently, these samples have been further characterized by sequence and cloning analysis (Calloway et al. manuscript in preparation).

To further characterize heteroplasmy, DNA from 131 bloodstains and 2554 head hairs from 132 individuals representing four population groups was isolated, amplified and typed using the LINEAR ARRAY mtDNA HVI/HVII Region-Sequence Typing Kit. In addition, morphological features of human head hair were characterized in order to further the understanding of the factors that influence amplification success rate and detection of heteroplasmy in hair tissue. The results of this study demonstrate differences in heteroplasmic expression between hair and blood tissue, higher levels in hair than in blood (Roberts and Calloway manuscript in preparation 1 and 2).

Conclusion

The discrimination power of the LINEAR ARRAY mtDNA HVI/HVII Region-Sequence Typing Kit approaches that of sequencing the HVI and HVII regions of the mitochondrial genome but the time, the cost, and the amount of sample consumed is substantially less. The availability of this rapid and simple PCR-based test for typing HVI/II regions of mtDNA has the potential to further expand the number and categories of forensic evidence samples that can be analyzed,

which in turn will increase the number of cases that can be resolved by crime laboratories at the federal, state, and local level.

Final Report

Introduction

The development over the past two decades of new technologies for the detection of genetic variation at the DNA level now allows us to use DNA profiling for individual identification. Today, numerous PCR-based DNA typing tests have been developed and are utilized for human identification in the analysis of biological evidence samples. PCR-based DNA typing tests targeting the nuclear genome (e.g. AmpliType® HLA DQA1 and PM, D1S80, STRs) have been particularly useful for individual identification because of their sensitivity and high discrimination power (Higuchi et al. 1988; Hochmeister et al. 1991; Comey et al.1993; Holt 2002). However, in some cases the analysis of nuclear genes fails because the template is limited or degraded. In these cases, polymorphisms within the mitochondrial genome can serve as a useful target. Today, an increased number of cases are submitted for mtDNA analysis and the number of laboratories performing mtDNA analysis continues to increase (Wilson et al. 1995; Allen et al. 1998; Carracedo et al. 2000; Melton and Nelson 2001). To help address these needs, we have developed and made commercially available a rapid method for analysis of sequence variation in the human mitochondrial genome that utilizes the established technologies of PCR amplification and immobilized probe hybridization.

Mitochondrial DNA

Mitochondria are organelles in the cellular cytoplasm, which are responsible for cellular respiration. Mitochondrial DNA is contained within the organelle and is haploid. The complete DNA sequence of the human mitochondrial genome was determined for one individual in 1981 (Anderson et al.) and hundreds of sequences have since been determined (Ingman et al. 2000;

Finnila et al. 2001; Herrnstadt et al. 2002). Mitochondrial DNA is a small, circular molecule of about 16,569 bp. The control region (or D-loop region) of mtDNA is an approximate 1000 bp region of noncoding DNA which contains one origin of replication and both origins of transcription as well as additional transcription and replication control elements. Mitochondrial DNA is highly polymorphic with the majority of the sequence variability concentrated in the control region, specifically Hypervariable regions I and II (Aquadro et al. 1993; Budowle et al. 1999; Miller et al. 2001; Imaizumi et al. 2002). The HVI and HVII regions are typically targeted for forensic identification purposes because of the high density of sequence variation.

Mitochondrial DNA has two additional unique features that make it particularly suitable for the analysis of biological remains, hairs, bone, teeth, and extremely limited or degraded DNA samples. First, mtDNA is inherited matrilineally (Giles et al. 1980). This mode of inheritance makes it a valuable genetic marker for the investigation of missing person cases because the subject's mother and siblings, as well as the mother's siblings (uncles and aunts), will all carry the same mtDNA sequence as that of the subject. Consequently, samples from maternally related individuals can be used as reference samples for the missing person.

Another unique feature of the mitochondrial genome is that it is present in high copy number. Alleles of the nuclear genes typed by the existing PCR-based tests are present in only one (spermatozoa and eggs) or two copies per cell, whereas mtDNA sequences can be present 100-1000 times per cell (Robin and Wong 1988). Since PCR amplification generally requires that the region of DNA containing the sequence of interest be intact (i.e., not degraded), the chances of obtaining a result from a severely degraded sample are significantly increased when mtDNA is

typed, simply because of its abundance relative to single copy nuclear DNA markers. The high copy number of the mitochondrial genome also frequently allows a result to be obtained from samples that have too little DNA to yield a result for a nuclear gene marker (e.g., telogen hairs).

Mitochondrial DNA Technology

Several approaches have been used to analyze the HVI/II regions of mtDNA for forensics: sequence based methods: direct DNA sequencing and pyrosequencing (Wrischnik et al. 1987; Sullivan et al. 1991; Ginther et al. 1992; Sullivan et al. 1992; Holland et al. 1993; Budowle et al. 1999) and probe based methods: conventional dot blot, Luminex hybridization and LINEAR ARRAY (reverse dot-blot) (Stoneking et al. 1991; Allen et al. 1998; Comas 1999; Reynolds et al. 2000; Gabriel, Calloway et al. 2001; Garbriel et al. 2003). Sequence analysis is currently the most widely used method. This procedure requires amplification of both HVI and HVII regions and is currently performed separately in two to four amplification reactions depending on the quality of DNA. If the DNA is extremely degraded, some laboratories take an approach to amplification and sequencing that has been used for studies of ancient DNA, which uses a series of primers that target very small overlapping segments of the HVI and HVII regions (Gabriel, Huffine et al. 2001). This strategy may involve as many as eight or more reactions and consumes an even greater amount of the often limited sample material.

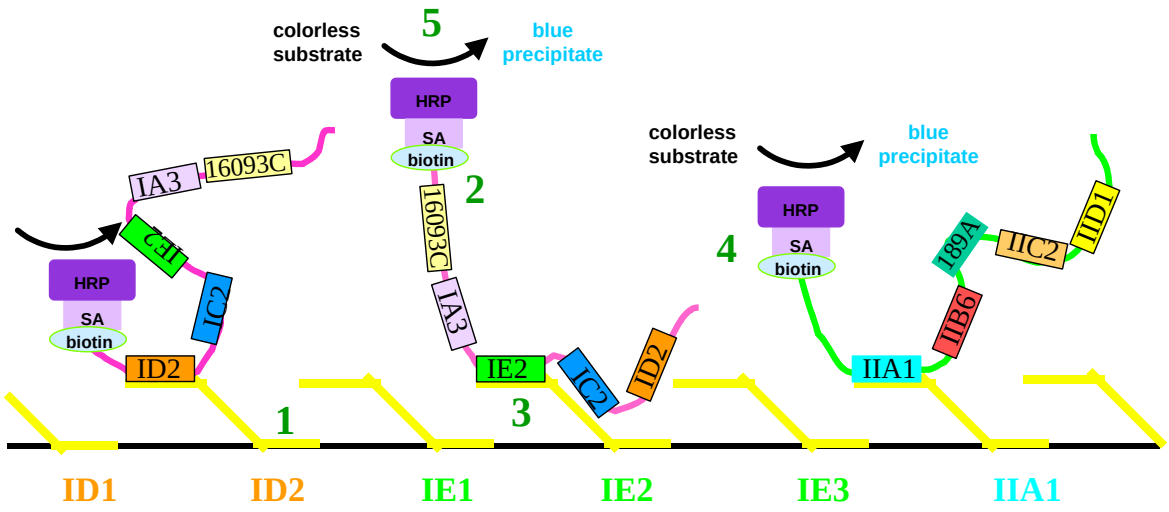
The first probe based mtDNA assay was developed with Dr. Mark Stoneking and has been used to screen large numbers of individuals in multiple populations (Stoneking et al. 1991; Melton et al. 1996; Melton et al. 1997a; Melton et al. 1997b; Melton et al. 2001). This conventional dot

blot assay consists of 23 SSO probes spanning nine regions within the two hypervariable regions of the mtDNA control region. Yet, when the number of probes greatly exceeds the number of samples, this typing procedure can become quite cumbersome because a membrane containing the fixed or immobilized PCR products has to be analyzed in a separate hybridization reaction with each labeled SSO probe. Thus, a genetic system with 23 probes would require 23 hybridization reactions and therefore would not be particularly amenable to routine forensic casework.

To overcome this problem, we developed a modification of this method in which SSO typing probes were immobilized in an array on a nylon membrane. This reverse hybridization approach means that the user does not have to perform an immobilization step and the results for all probes for a single individual are present in a single strip. The PCR product, which had been labeled with biotinylated primers during the amplification reaction, could be analyzed with all the probes in a single hybridization reaction (Saiki et al. 1989). The technical challenge for this format is to develop a set of probes that hybridize specifically under a single set of reaction conditions. In the immobilized probe format (known as the "reverse dot blot" or linear array), the binding of the labeled PCR product to a given probe in the immobilized probe array on the membrane could be detected using streptavidin (SA) conjugated to the enzyme horseradish peroxidase (HRP). Streptavidin has a very high affinity for biotin and HRP is capable of converting a soluble colorless chromogenic substrate such as tetramethylbenzidine (TMB) into a blue precipitate (Figure 1). The genotype or haplotype of the sample could then be determined from the pattern of blue dots or lines on the nylon membrane strip. To further simplify this typing procedure for some tests, we have developed software to interpret the pattern of probe reactivity as a genotype

(Bugawan et al. 1994) and developed prototype scanners to enter the probe reactivity pattern into the computer.

Figure 1.



1) **Probe covalently attached to nylon membrane**
2) **DNA template amplified with Biotinylated primer**
3) **Biotinylated product hybridizes to complementary probe**
4) **Streptavidin Horseradish peroxidase binds to biotin**
5) **Bound probes visualized by a color reaction**

The reverse dot blot method is the basis for the first commercial PCR genetic typing test, the AmpliType HLA DQα kit, which was introduced in early 1990 for use on forensic casework samples and as well as the AmpliType PM introduced late 1993. Since the development of these first assays, further improvements to the technology have been made including immobilization of the probes in lines instead of dots to allow for an increased number of probes per array. In addition to these commercially available forensics tests, we have also developed both research

and commercially available immobilized probe tests (linear arrays) for HLA class II (DRB1, DQB1, and DPB1) and class I (A, B, and C) loci tissue typing for transplantation (Erlich et al. 1991) and for detecting the mutations in the CFTR (cystic fibrosis) locus (CF Gold, Roche Diagnostics).

Development of the LINEAR ARRAY mtDNA HVI/HVII Region-Sequence Typing Kit

Our goal was to develop a sensitive, robust amplification and typing system to detect most of the diversity revealed by the HVI and HVII regions of the control region, but at a fraction of the time, effort, and cost of actually sequencing these regions. We chose to develop a mtDNA typing system that uses a set of SSO probes immobilized in lines on a strip of nylon membrane, since this method of typing is rapid, results are easy to interpret, and linear array probe panels are easily manufactured. Also, this method for typing mtDNA sequence variation can be performed in a single day using established DNA typing technologies and it can be used to quickly generate large population databases. Another advantage to the linear array technology is that it requires only a thermalcycler and a rotating water bath and can be implemented by any laboratory since it is simple to adopt and provides a low-budget solution to mtDNA typing compared to DNA sequencing. For high-throughput laboratories, automated instrumentation can be used to improve throughput. In addition, significant advances in automating the 'reading' of the linear arrays have been made.

Over the course of the granting period, we developed several mtDNA typing systems including a HVII linear array with a panel of 17 probes (Reynolds et. al. 2000; Calloway et. al. 2000), a HVI/HVII linear array with a panel of 27 probes (Gabriel et. al. 2001; Gabriel et. al. 2003), and the commercially available LINEAR ARRAY mtDNA HVI/II Region-Sequence Typing Kit

(Roche Applied Science, Indianapolis, IN) with a panel of 33 probes (Kline et. al. 2005; Divne et al. 2005). We describe below the various versions of the assay leading up to the final, commercially available system.

The mtDNA standard dot blot system developed with Dr. Mark Stoneking was used as a starting point for the mtDNA reverse hybridization assays. While a high percentage of the sequence diversity within populations is detectable using this set of 23 probes, data from screening mtDNA control regions from five different populations suggested that the discrimination power of the 23 probe standard dot blot system could be enhanced by addition of new probes. Specifically, adding additional probes targeting undetected variants in regions for which a high frequency of "blanks" (no probe signal in a given region resulting from a highly destabilizing mismatch within the probe binding region) are observed would increase the discrimination power. Because a blank provides information useful for typing but does not positively identify the actual sequence in a particular probe binding region, we found it desirable to add one or more probes that are able to hybridize to the undetected variants and, therefore, produce a positive probe hybridization result. Also, adding probes to target polymorphism outside the regions targeted by Stoneking et al. (1991) would increase the discrimination power. Many of the sequence variants targeted by this 23 probe system are targeted with the commercial LINEAR ARRAY probe panel, although a majority of the probes had to be redesigned to work optimally together with the same hybridization conditions necessary for the immobilized probe method.

The first 'reverse line blot' assay (linear array) for mitochondrial DNA sequence analysis was comprised of 17 SSO probes targeting the HVII region of the control region. The HVII region

was chosen over the HVI region as a target for the first assay since it had the greater potential for a higher discrimination power versus the HVI region. To decrease the number of blanks observed with the Stoneking assay and to further increase the power of discrimination for the HVII region, several additional probes were incorporated into the HVII probe panel including four probes detecting sequence variants in the B region and two probes detecting sequence variants in the C region. Sequence variation in 689 individuals from four population groups was determined using this HVII probe panel (Reynolds et al. 2000). The additional probes significantly decreased the observed number of blanks, increasing the power of discrimination compared to the original set of HVII probes used with the Stoneking assay. However, the discrimination power of the HVII probe panel was less than the 23 SSO probes targeting both HVI and HVII regions.

To increase the power of this assay, we incorporated a primer pair for the HVI region into our PCR amplification reaction and initially added probes to detect sequence variation in two regions of HVI (IA and IC) and at two heteroplasmy hotspots (16093 and 189). Probes detecting 189A/195C and 189A/195T are striped in a single line as well as probes detecting 189G/195C and 189G/195T to conserve space on the array and since the IIC region probes detect variation at 195. The "C-stretch" probe that detected length variation in the poly-C tract and the intensity control probe from the HVII array were removed. The C-stretch probe was removed because of the high frequency of length heteroplasmy in this region as well as the potential for cross-hybridization if HVI length heteroplasmy is present in the sample. This intermediary version of the HVI/HVII linear array was developed and used to establish a population database of 105 Croatians (Gabriel et al. 2001) and to identify victims of mass graves in Croatia (Gabriel et al. 2003).

To further increase the discrimination power of the assay, probes targeting two more regions of the HVI region were added, ID as defined by Stoneking et al (1991) and IE which was identified through sequence analysis. Probes targeting 16270 and 16278 (region IE) were chosen over probes targeting region IB, because the addition of the IE probes increased the discrimination power among Caucasians (for which the power of discrimination was among the lowest of the population groups tested). This version consists of 33 probes immobilized in 31 lines for detection of sequence variation at 18 positions spanning both hypervariable regions and is commercially available through Roche Applied Science as the LINEAR ARRAY mtDNA HVI/HVII Region-Sequence Typing Kit. These sites and probe designations are described in Table 1.

Table 1.

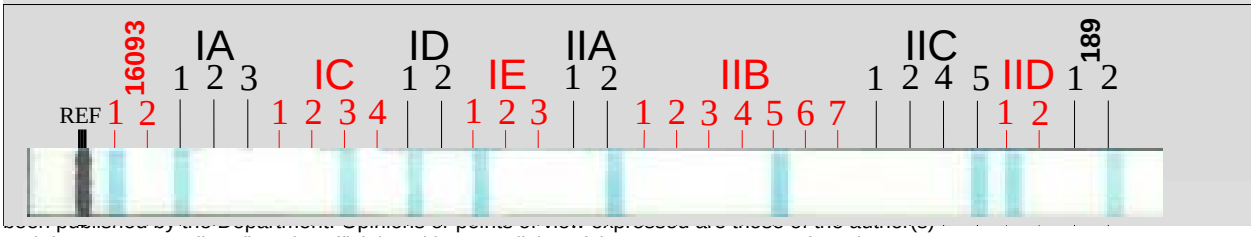
HVI Probe Designations	**Sequence Variation Detected**
	16093
16093 1	A T **T** T C
16093 2	· · **C** · ·
	16126 **16129**
IA1	T G **T** A C **G** G T
IA2	· · **C** · · · · ·
IA3	· · · · · **A** · ·
	16304 **16309** **16311**
IC1	A G **T** A C A T **A** G **T** A C
IC2	· · **C** · · · · · · · · · ·
IC3	· · · · · · · · · **C** · ·
IC4	· · · · · · · **G** · · · ·
ICw2/w3	· · **C** · · · · · · **C** · ·
	16362
ID1	C G **T** C C
ID2	· · **C** · ·
	16270 **16278**
IE1	C A **C** T A G G A T A **C** C A
IE2	· · · · · · · · · · **T** · ·
IE3	· · **T** · · · · · · · · · ·

Table 1 (continued)

HVII Probe Designations	Sequence Variation Detected
	73
IIA 1	G T **A** G T
IIA 2	· · **G** · ·
	146 **150** **152**
IIB1	C C **T** C A T **C** C **T** A T
IIB2	· · **C** · · · · · · · ·
IIB3	· · · · · · · · **C** · ·
IIB4	· · **C** · · · · · **C** · ·
IIB5	· · · · · · **T** · · · ·
IIB6	· · · · · · **T** · **C** · ·
IIB7	· · **C** · · · **T** · **C** · ·
	189 **195** **198** **200**
IIC1	G A **A** C A T A C **T** T A **C** T **A** A A
IIC2	· · · · · · · · **C** · · · · · · ·
IIC4	· · · · · · · · **C** · · **T** · · · ·
IIC5	· · **G** · · · · · · · · · · **G** · ·
	247
IID1	T T **G** A A
IID2	· · **A** · ·
	189 **195**
189 1	G A **A** C A T A C T T A
	· · · · · · · · C · ·
189 2	· · **G** · · · · · . · ·
	· · **G** · · · · · C · ·

The mtDNA linear array typing results for one individual (with mitotype 1131125512) are shown below in Figure 2. One of four categories of probe signal within each probe binding region is expected: (1) a single probe is positive, (2) a single probe signal is visible but its intensity is weaker than a positive signal in other regions within the same hypervariable region, (3) no probe signals are visible ("blank") resulting from a highly destabilizing mismatch within the probe binding region, or (4) two probe signals are visible.

Figure 2.



The mtDNA HVI/II linear array can be used as a screening tool to eliminate suspects quickly and to identify samples requiring further genetic analysis (i.e. sequencing). For example, suspects and specimens (e.g., multiple hairs) associated with a specific case can be rapidly screened with immobilized SSO probes. Samples from suspects who cannot be excluded and specimens that match the reference sample by the immobilized probe linear array could then be sequenced if desired without having to generate additional PCR product. Unlike standard amplification procedures for mtDNA sequencing analysis, the HVI/II regions are co-amplified, thereby reducing the amount of DNA material consumed during amplification by 50 to 75% as well as PCR set-up time. In addition, less than 50% (~20 of the 50 μL) of the PCR product generated with the assay is consumed during the quantification and hybridization steps, leaving the remainder of the PCR product (enough for forward and reverse sequencing of both HVI and HVII regions) for subsequent sequence analysis if needed. High quality DNA sequence has been consistently obtained using PCR product generated from the duplex reaction (Date Dhong et al. 2005). The very low to no background noise in the sequencing chromatogram is likely a result of the high specificity of the optimized duplex amplification reaction. This PCR system has been optimized to ensure both products amplify with approximately equal efficiency, primer dimer formation is minimized, and non-specific products are rarely generated. A higher amount of TaqGold DNA polymerase is used in the amplification to increase sensitivity and reduce the effect of potential PCR inhibitors.

Probes were optimized to minimize cross-hybridization, increase sensitivity and give approximately equal signal intensity within a hypervariable region. Typing conditions were

optimized for both a rotating water bath for manual typing and the Tecan Profiblot for automated typing. Three lines of commercial instrumentation, all reasonably inexpensive (~$20,000), for automating the hybridization, wash, and color development for the immobilized probe arrays are available (Profiblot, AutoReli or BeeBlot). To improve throughput of data analysis, software and hardware for scanning the developed "strips" and inferring a genotype are also available for the commercially available CF and HLA tests. Some modifications of the existing scanning and interpretation software will be required for the mtDNA typing system and are underway. Thus, screening samples with the linear array probe panel manually or using automated instrumentation could provide useful information within a day and decrease the number of specimens that need to be sequenced.

Developmental Validation

Before a new DNA typing system can be used by the forensic community, it must undergo developmental validation. This involves optimizing the test and evaluating its performance to establish its capabilities. Several improvements to the amplification and typing assay were made during optimization, including shortening the HVII reverse primer to further minimize primer dimer formation and using a single buffer for both hybridization and wash steps during typing. Once the systems were optimized and finalized, experiments designed to identify acceptable ranges in reaction parameters ('window studies') were completed. The parameters tested for the typing assay include stringency by varying temperature and salt conditions; formulation by varying concentrations of SDS, citrate, conjugation solution, and chromogen solution; varied time points for hybridization, stringent wash, color development, conjugation, and water washes; and varied volumes of PCR product. The PCR parameters tested include time and temperature

varied for annealing, activation, denaturation, extension, and final extension; formulation by varying concentrations of magnesium chloride, Taq Gold, and dNTPs; and primer concentrations. Equivalent results were observed for the specified range for each parameter tested necessary for manufacturing allowances.

Additionally, a series of studies were conducted to evaluate the performance of the assay as suggested by the Technical Working Group on DNA Analysis Methods (TWGDAM) committee and DNA Advisory Board (DAB). Results from a subset of these studies are summarized below and will be submitted for publication. Studies to evaluate sensitivity, degradation, reproducibility and detection of mixtures were also conducted at the California Department of Justice Richmond Laboratory and are also summarized here.

Species Specificity

Species specificity of the mtDNA LINEAR ARRAY Amplification and Typing system was examined using nonhuman DNA extracts. Three concentrations (500, 50, and 5 pg) of DNA samples from 16 species were amplified and typed: primates (gorilla, chimpanzee, orangutan, marmoset) and non-primates (mouse, rat, dog, cat, fish, cow, sheep, chicken, turkey, pig, e-coli, candida). PCR products were generated for higher order primates (gorilla, chimpanzee, and orangutan) and rat, detected by gel electrophoresis. The concentration of PCR product generated from the rat DNA was very low, <2.5 ng/ul for the highest input of DNA. The size of the product generated from the rat DNA was ~350 bp, much smaller than the size of human HVI/HVII products. All samples were typed with the linear array probe panels. Only products

for higher order primates hybridized and only to a subset of probes in the panel. No probe signals were visible for non-primates, including rat.

To further test interference of nonhuman DNA with human DNA, 500 pg of each of the 16 nonhuman DNAs were mixed with 500, 50 and 5 pg of human DNA and amplified and typed. In all cases, the yield of the PCR product generated from the human DNA as visualized by gel electrophoresis was not affected by the presence of the tested nonhuman DNAs. Only the expected human type was observed with the linear array probe panels for all non-primates, including rat. A mixture of the higher primate type and human type were observed for the chimpanzee, gorilla, and orangutan.

Sensitivity

Optimal intensity of probe signals on the linear arrays was observed when approximately 1.0 pg or more of total genomic DNA was amplified for 34 cycles. Typing of less than 0.5 pg of DNA sample required 38 cycles of amplification to obtain enough amplicon for optimal signal with the array. Stochastic effects were observed for some samples that were amplified for 38 cycles with 0.05 pg of template DNA. Studies comparing duplex and single-plex reaction mixtures showed that amplifying the HVI and HVII regions in duplex produced linear array results that were comparable to those obtained when amplifying each region separately.

Degradation

A time course digestion of template DNA with DnaseI enzyme resulted in DNA fragments with varied average lengths. Typing results showed no loss of signal for DNA fragments of average

length of 500 base pairs or more at amplification of 34 cycles. DNA fragments with average lengths of 500 base pairs or less required amplification for 38 cycles for optimal typing results.

Reproducibility

DNA extracted from bloodstains, hairs, buccal swabs, nail clippings, and non-probative samples showed reproducible mtDNA types when tested multiple times. Identical typing results were obtained from thirteen hair samples from one individual with the exception of one hair, which displayed apparent heteroplasmy. The hair samples were also subjected to DNA sequence analysis, which confirmed a position change for the one hair that displayed heteroplasmy with the linear array probe panel .

Mixture Studies

Two DNA samples that differ at six positions were mixed in defined ratios and typed. The results indicated that the minor component was detectable down to a 10:1 ratio, the lowest ratio tested in this study. Additional mixtures studies have been conducted and in some cases the minor component was detectable down to a 20:1 ratio.

Population Studies

A series of population studies was conducted to determine the distribution of the various mitotypes and to calculate the level of discrimination power for the linear array probe panel for a number of population groups. Data from a subset of these studies are summarized below.

Detection of Sequence Variation in the HVII Region of the Human Mitochondrial Genome in 689 Individuals Using Immobilized Sequence-Specific Oligonucleotide Probes

Using a panel of 17 SSO probes immobilized on nylon membrane strips targeting the HVII region, blood samples from 689 individuals from four US population groups were typed. These results are published in the Journal of Forensic Science (Reynolds et al. 2000). The genetic diversity value for each population was calculated from the frequency data collected for the HVII probe panel and range from ~0.95 to 0.98, highest among African Americans. The frequencies of each distinct mtDNA type in each group were determined (see Table 3 in Reynolds et al. 2000). DNA sequence analysis was performed on a subset of the samples to characterize the sequences associated with the blanks and weak probe signals.

Detection of Sequence Variation in the HVI and HVII regions of 674 Individuals

To test the final version of the HVI/HII assay and to generate a large population database, 674 available samples typed previously with the HVII array were amplified with the final primer and premix and typed with the final version of the HVI/HVII linear array. This database was used to calculate a power of discrimination for each of the four populations to determine the mtDNA type probability that a random individual would have the same mtDNA type as the evidence sample. The genetic diversity values ranged from ~9.5-9.9 (see Table 2).

Table 2.

Genetic diversity value determined for four US populations using the LINEAR ARRAY mtDNA HVI/HVII Region-Sequence Typing Kit.

Population Group	# of samples analyzed	# different mitotypes	# unique mitotypes	h value
African American	194	137	111	0.993
U.S. Caucasian	197	99	73	0.977
U.S. Hispanic	197	91	67	0.945
Japanese	86	58	48	0.981

The highest discrimination power was observed for the African Americans and the lowest discrimination power was observed for the Hispanic and Caucasians groups. A large number of individuals having a few common types account for the lower discrimination power in these two populations. The most common Caucasian mitotype (1111111111) occurred 23 times in 198 samples (~10%). Approximately 5% of individuals (or 50% of those with this mitotype) still shared the same HVI/II sequence (263G, 315.1C). This sequence haplotype (263G, 315.1C) has been identified as the most frequently observed Caucasian HV sequence and was observed in ~7% of the population based on the international mtDNA sequence database (Parsons and Coble 2001). The second most common mitotype (11111131111) occurred nine times with six of these nine individuals sharing the same HVI/II sequence. Also, two common mitotypes were found in ~30% of the 197 US Hispanics tested and sequence analysis had a limited ability to further distinguish individuals with one of the two common mitotypes (1111211011). Thus, the greatest

limitation of both HVI/II sequence *and* linear array analyses lies with the power of discrimination due to the small number of common types. Variation outside the HV regions will need to be targeted to further discriminate individuals having these common types. Although several common mitotypes were observed, most of the mitotypes occurred only once within the population group (see # unique mitotypes in Table 2). This data has been presented at several conferences and a manuscript is in preparation (Calloway et al.).

In addition, the HVI and HVII regions for each of the 674 samples was sequenced to verify the typing results. This sequence data was used to characterize samples with 'weak' signals and 'no' probe signal ('blank') within a given region and to determine the sequence variation underlying the probe causing the mismatch. Also, a power of discrimination will be calculated for the sequence database for this population and compared to the discrimination power from the linear array database.

Population variation of human mtDNA hypervariable regions I and II for 105 Croatian individuals by immobilized sequence-specific oligonucleotide (SSO) probe analysis

Sequence variation detection for 105 randomly selected Croatian individuals with an array of 27 immobilized SSO probes that targets six regions within HVI and HVII and two additional sites (189 and 16093) was used to establish a reference database for mtDNA analysis of human remains recovered from mass graves. Results from this population study using an early version of the HVI/HVII mtDNA assay are published (Gabriel et al. 2001). For this database, 50 different mitotypes were observed, among which 33 were unique. The corresponding haplotype gene diversity value (h) for this database is ~0.952 using this 27 linear array probe panel. The

gene diversity value would be slightly higher using the current panel of 33 probes for this population. Discrimination power is reduced in this population because of the high frequency of two common mitotypes. The most frequent mitotypes occurred 18 times or ~17.1% (111111 189 (A) 16093 (T)) and 11 times or ~10.5% (131111 189 (A) 16093 (T)); all other mitotypes occurred 5% or less using this 27 probe panel. Direct sequence analysis was performed for samples with the two most frequent mitotypes to determine the presence of additional polymorphic sites that might offer increased discrimination to the linear array analysis. Of the 18 samples with the most frequent mitotype, one set of five samples and two sets of two samples could not be distinguished by direct sequence analysis of the HVI and HVII regions. Furthermore, sequence analysis could not distinguish the 11 samples with the second most frequent mitotype, since no additional polymorphisms were observed in either the HVI or HVII region. Thus, further discrimination of these samples would require targeting areas outside the HV regions as suggested above. The population variation described here was successfully applied to a separate study summarized below, which analyzed 18 skeletal elements recovered from a Croatian mass grave (Gabriel et al. 2003).

Detection of Sequence Variation in the HVI and HVII Region of 200 Individuals Residing in Georgia

The current HVI/HVII linear array was also used to generate a regional database for Georgia at the Georgia Bureau of Investigation (GBI). For this database, blood samples randomly collected from 100 Caucasians and 100 African Americans from individuals who resided in Georgia, previously used to generate their regional STR and AmpliType PM databases, were typed with the SSO linear array. The HVI and HVII regions of each of the 200 samples were sequenced as

well. The genetic diversity value for each population was calculated from the frequency data and compared for both (sequencing and linear array) typing methods. We also estimated the frequency of heteroplasmy detected by both typing methods. These findings are reported in a manuscript in preparation (Williams et al. manuscript in preparation).

Detection of Sequence Variation of 666 Individuals using an Automated Typing Instrument

The LINEAR ARRAY™ mtDNA HVI/HVII Region-Sequencing Typing Kit (Roche Applied Science, Indianapolis, IN) was used to type 666 individuals from U.S. Caucasian, African American, and Hispanic groups at the National Institute of Standards and Technology Laboratory. Processing of the LINEAR ARRAY probe panels was automated on a Tecan ProfiBlot workstation. Observable variation in 666 individuals is reported and frequencies of the mitotypes within and between populations are presented in a manuscript submitted for publication (Kline et al. 2005). Samples exhibiting the most common Caucasian mitotype were subdivided with a multiplexed amplification and detection assay using eleven single nucleotide polymorphisms in the mitochondrial genome.

Beta Site Trials and Analysis of Case type samples

Throughout the course of development, this assay was extensively evaluated within our laboratory as well as nearly 20 laboratories who participated in beta studies or collaborative projects. Results from a subset of these studies are summarized below.

Beta Site Trials

The Beta site trials were designed as a three part training study followed by an independent study. For study1A the participant was provided 11 DNA samples and PCR products derived from those samples. Participants were asked to amplify the DNAs, compare the yields on a gel to the provided PCR products, and type all PCR products with the linear arrays. The goal of this part of the study was to ensure the participant could amplify DNA without introducing any contaminants and successfully type and interpret the arrays. For study1B the participant was provided 15 DNA samples varying in concentration and including some samples with mixtures. The goal of this study was to demonstrate the participant could successfully quantitate the amount of PCR product, dilute the product or increase the yield by further amplification cycles if necessary, and successfully type and interpret the arrays. This study included several heteroplasmic samples as well as samples which were intentionally mixed in order to be a challenging interpretation test. For study 1C the participant was provided with two hairs from a single individual and was asked to extract the hairs, amplify the DNA, and compare the yield and type to a provided extract from the same individual. The study was designed to test the participants' ability to successfully extract hairs without introducing contamination. For further practice, the participant was asked to extract five hairs from two additional individuals.

Results from the initial beta studies were generally as expected, however several laboratories consistently had weaker than expected probe signal intensities and imbalanced signals within a probe panel. A higher than optimum pH of the Wash Buffer used in the hybridization and wash steps was identified as the cause of these unexpected results after a series of troubleshooting experiments. In our initial protocol, we did not include a step to check the pH of the final Wash

buffer. However during the beta studies, we discovered that variation in the pH of the 2X SSPE, 0.5% SDS Wash Buffer from the optimum 7.4 can impact probe signal intensities. Results from our 'window studies' showed that if the pH of the Wash Buffer is eight or greater, probe signal intensities may be weaker than expected and imbalanced. Results from the Beta site studies indicate that the pH of DI water can vary between laboratories and can impact the final pH of the Wash Buffer. Variation from the recommended starting pH of the SSPE and SDS can lead to variation of the pH of the Wash Buffer. To resolve this issue, we have now implemented a step whereby the laboratory is asked to adjust the pH of the final Wash Buffer to 7.4 (±0.2) in addition to the individual reagents.

The initial beta site studies led to several modifications to improve the amplification and linear array assay. To validate the final version of the assay, new reagents and materials for studies 1A and 1B were sent to a subset of the participating laboratories and this part of the study was repeated.

Identification of Human Remains by Immobilized Sequence-specific Oligonucleotide Probe Analysis of mtDNA Hypervariable Regions I and II

In collaboration with Dragan Primorac of Croatia, we analyzed the mtDNA control region of 18 skeletal elements recovered from a mass grave in Croatia that could not be typed using nuclear markers (PM+DQA1 and short tandem repeats); the results have been published recently (Gabriel et al 2003). For 14 of 18 sample extracts, duplex PCR amplification of the mtDNA hypervariable regions I and II (HVI and HVII) (444bp and 415bp amplicons, respectively) was successful in providing sufficient product for typing with an array of immobilized sequence-

specific oligonucleotide (SSO) probes. One set of remains was positively identified using a panel of immobilized SSO probes that targets six polymorphic regions and two additional sites within the human mtDNA HVI and HVII. The corresponding mitotype obtained for the bone sample and the putative maternal reference was unique in a database of 105 randomly selected Croatian individuals. Following immobilized SSO probe testing, full sequence analysis of the HV regions was performed to confirm the match.

Forensic Casework Analysis Using the HVI/HVII mtDNA LINEAR ARRAYs.

Over the past five years Marie Allen and her colleagues at Uppsala University, Sweden have used various versions of the mtDNA immobilized SSO probe assay on more than 300 forensic samples. Most recently, they have evaluated the final LINEAR ARRAY mtDNA HVI/HVII Region-Sequence Typing Kit in 16 forensic cases, comprising 89 samples. Using the HVI/HVII linear array probe panel, 57% of the samples were excluded and thus only 43% of the samples (38 out of 89 samples) required further sequencing due to a match or inconclusive results. Of the 38 samples submitted for sequencing, 12 samples were excluded because of additional variation detected by sequence analysis. Thus, 81% of the samples (51 out of 63 samples) that differed in the HVI and HVII sequence from the reference sample could be excluded using the HVI/HVII linear array probe panel alone.

Using the LINEAR ARRAY HVI/HVII Region-Sequence Typing kit, Allen and colleagues demonstrate the possibility to decrease sequencing efforts substantially and thereby increase the turn-around time in casework analysis. A large number of crime scene samples can be screened simultaneously for inclusion or exclusion with respect to reference samples and thereby identify

the samples of most interest for further investigation. Furthermore, as the PCR is performed in a duplex reaction and remaining PCR-products can be sequenced directly, valuable evidence material is saved. Overall, Allen and her colleagues report that the linear HVI-HVII linear array assay is a robust, rapid, accurate and sensitive method with a high potential to discriminate between different mtDNA types. The use of the mtDNA linear arrays in their laboratory has served as a valuable pre-screening method and demonstrates a potential to reduce the sequencing efforts by more than half. A manuscript reporting these findings is in press (Divne et al. 2005).

Analysis of Adjudicated Case Samples Using the HVI/HVII Linear Array Probe Panel

Samples from multiple adjudicated cases were typed at the Georgia Bureau of Investigation Laboratory using the SSO linear array and sequence analysis. Cases in which the suspect had been excluded by STR typing were chosen for this study to allow us to assess the value of the linear array assay as a screening tool for the exclusion of individuals. In all but one case, linear array typing was sufficient to exclude the suspects who had been excluded by STR analysis. In this particular case, the suspect excluded by STR analysis had the same SSO mitotype as well as the same HVI and HVII sequence as the donor of the semen stain. Prior to mtDNA typing, it was thought that the suspect was a brother of the donor of the semen stain based on STR analysis. The mtDNA analysis was consistent with this possibility.

Mock Case Study

To investigate the value of the linear arrays as a screening tool, reference hairs and blood samples from victims and suspects were obtained from rape kits collected for sexual assault cases. These samples were assembled into eight mock cases such that each was comprised of

three questioned pubic hairs, a bloodstain from one of the victims, and a reference bloodstain from one of the suspects, not necessarily from the same adjudicated case. The mitochondrial DNA sequences of the samples were not known prior to assembling the mock cases. Two laboratories extracted and analyzed the samples (Georgia Bureau of Investigation and San Bernardino Sheriff's Office). The linear array first used to type these samples was comprised of the full panel of 33 probes, however modifications to a subset of the probe sequences were subsequently made for the final version of the assay. For this study, all samples were sequenced and all of the typing results between the two laboratories were compared. In all cases, the correct associations between questioned and reference samples were made. Samples were typed a second time using the final version of the assay by one laboratory (San Bernardino Sheriff's Office). Results are summarized by Mehul Anjaria as part of his Master's Thesis (Anjaria 2003).

Samples collected from Different Media and Subjected to Various Environmental Insults

In collaboration with the Mehul Anjaria at the San Bernardino County Sheriff's Department, this study was undertaken to assess the ability to obtain results from DNA recovered from biological samples deposited on various substrates and subjected to environmental and chemical insults. The samples chosen for analysis were bloodstains collected at actual crime scenes in San Bernardino County, CA sometime prior to 1993. The samples are considered 'secondary reference samples', meaning that their source can be logically inferred (e.g. blood collected from a pool adjacent to a body with a gunshot wound to the head). These samples had been exposed to an array of environmental conditions (e.g. snow, heat, rain) and were deposited on a wide variety

of substrates (e.g. carpet, clothing, asphalt, dirt). Also analyzed were the actual reference (origin positively known) blood samples.

Existing DNA extracts from these samples produced from either the organic (phenol/chloroform) method or the Chelex method were used. These samples were not extracted with mitochondrial DNA analysis in mind. Also, various analysts performed these extractions in either 1993 or 2000 and some samples were extracted multiple times.

Using early versions of the HVI/HVII linear array probe panel, all but three of the crime scene samples were successfully amplified and typed following the appropriate protocols. For one sample, no additional extract remained for further testing. It was determined that the two remaining non-amplifying samples likely contained substances inhibitory to the PCR. Varying parameters such as amount of input DNA, using alternate primer sets, and the use of bovine serum albumin were employed in an effort to overcome the inhibition. Successful amplification was achieved for both samples simply by decreasing the volume of the input DNA into the PCR. No contamination was observed when comparing crime scene samples with the corresponding reference samples.

Characterization of Heteroplasmy

Heteroplasmy holds implications for forensic analysis of specimens and needs to be considered to ensure that mtDNA typing results are interpreted appropriately. In forensic analysis, the presence of heteroplasmy impacts the interpretation of mtDNA matches between evidentiary materials and individuals or their maternal relatives (Melton 2004). Heteroplasmy is most often

detected in the control region as length variation within the homopolymeric tract of cytosine residues in the HVI and HVII regions, referred to as length heteroplasmy (Hauswirth and Clayton 1985; Bendall and Sykes 1995; Bendall et al. 1996; Marchington et al. 1997). Heteroplasmic point mutations were once thought not to occur in the normal population (Monnat and Loeb 1985; Monnat and Reay 1986) but there is increasing evidence that the frequency of these point mutations in the normal population is significant (Gill et al. 1994; Comas et al. 1995; Jazin et al. 1996; Calloway et al. 2000; Tully et al. 2000). Therefore, it is important for a mtDNA practitioner to have a thorough understanding of the biological underpinnings of heteroplasmy and its frequency within each forensically important sample type (Melton 2004).

For these reasons, heteroplasmy was extensively studied in our laboratory. In the first study, 459 hairs from 23 people were analyzed using the HVII linear array. Point heteroplasmy was observed in one or more hairs in 22% of individuals. In contrast, only 0.8% of the 689 blood samples from individuals typed using the HVII array to generate a population data base were heteroplasmic (point) (Reynolds et al. 2000). Results from additional studies to characterize heteroplasmy are summarized below.

Characterization of Heteroplasmy among Various Tissue Samples

To investigate the frequency of heteroplasmy across various tissues and age groups, the HVII array was used to screen for heteroplasmy from heart, brain, muscle, and blood samples from 43 cadavers (Calloway 1998; Calloway et al. 2000). Results showed (a) higher levels in muscle tissue than in the other tissues, blood showing the least, and (b) higher levels in tissues from older individuals (Calloway 1998; Calloway et al. 2000).

Sequence analysis conducted to confirm heteroplasmy in the five samples identified by the HVII linear array probe panel revealed additional heteroplasmic sites outside the initially targeted areas. Heteroplasmy was observed at the same site in 4 of the 5 muscle tissue samples sequenced, suggesting the possibility of tissue specific heteroplasmic "hot spots" (specifically at position 189, probes for which were later added to the linear array probe panel). To further characterize heteroplasmy, sequence analysis of the samples from the 43 cadavers was conducted and revealed a high degree of site heteroplasmy in the HVII region outside the regions detected by the immobilized SSO probe assay. Very little heteroplasmy was observed in the HVI region. Heteroplasmy was observed at some positions more frequently than others (sites 72, 73, 185, 189), suggesting heteroplasmic hotspots. Also consistent with our initial observations, the frequency of heteroplasmy differed across tissue types and was highest in muscle tissue. At some sites, heteroplasmy was only observed in muscle tissue, while at other sites heteroplasmy was observed in all tissues but blood. Since the mtDNA type was not found to be identical from tissue to tissue and differed across age groups, the observations from this study should be considered if mtDNA typing is used for identification of remains from mass disaster (i.e. chose tissue other than muscle if available).

Analysis of Human Head Hairs and Bloodstains from 132 Individuals

Katherine Roberts at California State, LA examined mitochondrial DNA polymorphisms in human head hair and bloodstains with respect to their potential for forensic application. The particular focus of this collaborative project was to characterize the morphological features of human head hair in order to further the understanding of the factors that influence amplification success rate and characterize heteroplasmy in hair tissue. DNA from 131 bloodstains and 2554

head hairs from 132 individuals representing four population groups was isolated, amplified and polymorphisms were detected using the mtDNA linear array assay. The hair samples were characterized according to their hair growth phase: 1303 were identified microscopically as telogen in origin and 1251 were classified as anagen. Amplification success was assessed as a function of several independent variables: morphological characteristics; hair growth phase; donor age and gender; scalp origin; use of cosmetic hair treatments; and race of the donor. A manuscript reporting these findings is in preparation (Roberts and Calloway manuscript 1 in preparation).

Samples that successfully amplified were typed using the HVI/HVII mitochondrial DNA linear array assay. The genetic diversity value for each population group was analyzed and the frequency of each mtDNA haplotype was determined. For the purpose of this study, a sample was scored as heteroplasmic if two probe signals were visible within a single probe region (either with equal or uneven intensity). The results of this study demonstrate differences in heteroplasmic expression between hair and blood tissues (hair>blood). Finally, the study evaluates the frequency of heteroplasmy across racial population groups and assesses whether the heteroplasmic condition differs significantly with age, gender, medulla morphology, region of the scalp, hair growth or, when comparing living and deceased donors. A second manuscript reporting these results is in preparation (Roberts and Calloway manuscript 2 in preparation).

Conclusion

Developmental validation of the LINEAR ARRAY mtDNA HVI/HVII Region-Sequence Typing Kit has been completed and internal validation within forensic DNA testing laboratories is on-

going. This mtDNA HVI/II amplification and linear array system and strategy has been used successfully in Sweden for casework samples for several years (Allen et al. 1998; Divne et al. 2005. This assay has proven to be an effective tool for analyzing the origin of hairs shed at a crime scene and reducing the number of forensic samples that require sequencing. The system also has been used successfully to identify remains from a mass grave site in Croatia (Gabriel, Calloway et al. 2001; Gabriel et al. 2003). The availability of this rapid and simple PCR-based test for typing HVI/II regions of mtDNA has the potential to further expand the number and categories of forensic evidence samples that can be analyzed, which in turn will increase the number of cases that can be resolved by crime laboratories at the federal, state, and local level.

References

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

Anjaria, Mehul B. (2003). "Evaluation of Prototypes of Roche Applied Science's LINEAR ARRAY Mitochondrial DNA HVI/HVII Region-Sequence Typing Kit for Forensic Mitochondrial DNA Typing. Master's Thesis. California State University, Los Angeles.

Allen, M., A. S. Engstrom, et al. (1998). "Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: sensitivity and matching probabilities." J Forensic Sci **43**(3): 453-64.

Aquadro, C. F. and B. D. Greenberg (1983). "Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals." Genetics **103**(2): 287-312.

Bendall, K. E. and B. C. Sykes (1995). "Length heteroplasmy in the first hypervariable segment of the human mtDNA control region [see comments]." Am J Hum Genet **57**(2): 248-56

Bendall, K. E., V. A. Macaulay, et al. (1996). "Heteroplasmic point mutations in the human mtDNA control region." Am J Hum Genet **59**(6): 1276-87.

Budowle, B., M. R. Wilson, et al. (1999). "Mitochondrial DNA regions HVI and HVII population data." Forensic Sci Int **103**(1): 23-35.

Bugawan, T.L., R. Apple and H.A. Erlich. 1994. A method for typing polymorphism at the HLA-A locus using PCR amplification and immobilized oligonucleotide probes. Tissue Antigens 44:137-147.

Calloway, Cassandra D.. The frequency of heteroplasmy in the HVII region of mtDNA in human tissues. Master's Thesis. University of Georgia, Athens, Georgia 1998.

Calloway, C. D., R. L. Reynolds, et al. (2000). "The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age." Am J Hum Genet **66**(4): 1384-97.

Carracedo, A., B. a. W, et al. (2000). "DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing." Forensic Sci Int **110**(2): 79-85.

Comas, D., S. Paabo, et al. (1995). "Heteroplasmy in the control region of human mitochondrial DNA." Genome Res **5**(1): 89-90.

Comas, D., R. Reynolds, et al. (1999). "Analysis of mtDNA HVRII in several human populations using an immobilised SSO probe hybridisation assay." Eur J Hum Genet **7**(4): 459-68.

Comey, C.T., B. Budowle, D.E. Adams, A.L. Baumstark, J.A. Lindsey and L.A. Presley. 1993. PCR amplification and typing of the HLADQα gene in forensic samples. Journal of ForensicSciences 38:239-249.

Date Chong, M., C.D. Calloway, S.B. Klein, C. Orrego, and M.R. Buoncristiani. Optimization and Validation of a DNA Sequencing Strategy for the Analysis of Mitochondrial DNA for Forensic Casework. Forensic Science International, 2005. In Press

Divne, Anna-Maria, Martina Nilsson, Cassandra Calloway, Rebecca Reynolds, Henry Erlich and Marie Allen. Forensic Casework mtDNA analysis using a Reverse HVI-HVII Sequence-Specific Oligonucleotide Probe Assay. J Forensic Sci, 2005. In Press

Erlich, H., T. Bugawan, A.B. Begovich, S. Scharf, R. Griffith, R. Saiki, R. Higuchi and P.S. Walsh. 1991. HLA-DR, DQ and DP typing using PCR amplification and immobilized probes. Eur. J. of Immunogen. 18:33-55.

Finnila, S., M.S. Lehtonen, and K. Majamaa, *Phylogenetic network for European mtDNA.* Am J Hum Genet, 2001. **68**(6): p. 1475-84.

Gabriel, M.N., et al., *Identification of human remains by immobilized sequence-specific oligonucleotide probe analysis of mtDNA hypervariable regions I and II.* Croat Med J, 2003. **44**(3): p. 293-8.

Gabriel, M. N., C. D. Calloway, et al. (2001). "Population variation of human mitochondrial DNA hypervariable regions I and II in 105 Croatian individuals demonstrated by immobilized sequence-specific oligonucleotide probe analysis." Croat Med J **42**(3): 328-35.

Gabriel, M. N., E. F. Huffine, et al. (2001). "Improved MtDNA sequence analysis of forensic remains using a "mini-primer set" amplification strategy." J Forensic Sci **46**(2): 247-53.

Giles, R.E., H. Blanc, H.M. Cann and D.C. Wallace. 1980. Maternal inheritance human mitochondrial DNA. Proc Natl. Acad. Sci. USA 77:6715-6719.

Gill, P., P. L. Ivanov, et al. (1994). "Identification of the remains of the Romanov family by DNA analysis." Nat Genet **6**(2): 130-5.

Hauswirth, W. W. and D. A. Clayton (1985). "Length heterogeneity of a conserved displacement-loop sequence in human mitochondrial DNA." Nucleic Acids Res **13**(22): 8093-104.

Herrnstadt, C., J. L. Elson, et al. (2002). "Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups." Am J Hum Genet **70**(5): 1152-71.

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

Hochmeister, M.N., B. Budowle, U.V. Borer, U. Eggmann, C.T. Comey and R. Dirnhofer. 1991. Typing of deoxyribonucleic acid (DNA) extracted from compact bone from human remains. Journal of Forensic Sciences 36:1649-1661.

Holland, M.M., D. L. Fisher, L.G. Mitchell, W.C. Rodriguez, J.J. Canik, C.R. Merrill and V.W. Weedn. 1993. Mitchondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam war. J. Forensic Sciences 38:542553.

Holt, C.L., et al., *TWGDAM validation of AmpFlSTR PCR amplification kits for forensic DNA casework.* J Forensic Sci, 2002. **47**(1): p. 66-96.

Imaizumi, K., T. J. Parsons, et al. (2002). "A new database of mitochondrial DNA hypervariable regions I and II sequences from 162 Japanese individuals." Int J Legal Med **116**(2): 68-73.

Ingman, M. and U. Gyllensten (2001). "Analysis of the complete human mtDNA genome: methodology and inferences for human evolution." J Hered **92**(6): 454-61.

Jazin, E. E., L. Cavelier, et al. (1996). "Human brain contains high levels of heteroplasmy in the noncoding regions of mitochondrial DNA." Proc Natl Acad Sci U S A **93**(22): 12382-7.

Kline, Margaret C., Peter M. Vallone, Janette W. Redman, David L. Duewer, Cassandra D. Calloway, and John M. Butler. Mitochondrial DNA Typing Screens with Control Region and Coding Region SNPs. J Forensic Sci, 2005. 50(2)

Marchington, D. R., G. M. Hartshorne, et al. (1997). "Homopolymeric tract heteroplasmy in mtDNA from tissues and single oocytes: support for a genetic bottleneck." Am J Hum Genet **60**(2): 408-16.

Melton, T. and M. Stoneking (1996). "Extent of heterogeneity in mitochondrial DNA of ethnic Asian populations." J Forensic Sci **41**(4): 591-602.

Melton, T., M. Wilson, et al. (1997a). "Extent of heterogeneity in mitochondrial DNA of European populations." J Forensic Sci **42**(3): 437-46.

Melton, T., C. Ginther, et al. (1997b). "Extent of heterogeneity in mitochondrial DNA of sub-Saharan African populations." J Forensic Sci **42**(4): 582-92.

Melton, T., S. Clifford, et al. (2001). "Diversity and heterogeneity in mitochondrial DNA of North American populations." J Forensic Sci **46**(1): 46-52.

Melton, T. and K. Nelson (2001). "Forensic mitochondrial DNA analysis: two years of commercial casework experience in the United States." Croat Med J **42**(3): 298-303.

Melton, T. (2005). "Mitochondrial DNA Heteroplasmy". Forensic Science Review 16 (1):1-20

Miller, K.W. and B. Budowle, *A compendium of human mitochondrial DNA control region: development of an international standard forensic database.* Croat Med J, 2001. **42**(3): p. 315-27.

Monnat, R. J., Jr. and L. A. Loeb (1985). "Nucleotide sequence preservation of human mitochondrial DNA." Proc Natl Acad Sci U S A **82**(9): 2895-9.

Monnat, R. J., Jr. and D. T. Reay (1986). "Nucleotide sequence identity of mitochondrial DNA from different human tissues." Gene **43**(3): 205-11.

Parsons, T. J. and M. D. Coble (2001). "Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial DNA genome." Croat Med J **42**(3): 304-9.

Reynolds, R., K. Walker, et al. (2000). "Detection of sequence variation in the HVII region of the human mitochondrial genome in 689 individuals using immobilized sequence-specific oligonucleotide probes." J Forensic Sci **45**(6): 1210-31.

Roberts, Katherine A. and Cassandra Calloway. Mitochondrial DNA Amplification Success Rate as a Function of Hair Morphology (in preparation).

Roberts, Katherine A. and Cassandra Calloway. Detection of Heteroplasmy in Human Head Hair and Bloodstains from 132 Individuals using a Mitochondrial DNA Linear Array Assay (in preparation)

Robin, E.D. and R. Wong. 1988. Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. J. Cell. Phys. 136:507-513.

Saiki, R.K., P.S. Walsh, C.H. Levenson and H.A. Erlich. 1989. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. Proc. Natl. Acad. Sci. USA 86:6230-6234.

Stoneking, M., D.Hedgecock, R.G. Higuchi, L. Vigilant and H.A. Erlich. 1991. Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. Am.J. Hum. Genet. 48:370-382.

Sullivan, K.M., R. Hopgood, B. Lang and P. Gill. 1991. Automated amplification and sequencing of human mitochondrial DNA. Electrophoresis 12:17-21.

Tully, L. A., T. J. Parsons, et al. (2000). "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 **67**(2): 432-43.

Wilson, M. R., J. A. DiZinno, et al. (1995). "Validation of mitochondrial DNA sequencing for forensic casework analysis." Int J Legal Med **108**(2): 68-74.

Wrischnik, L.A., R.G. Higuchi, M, Stoneking, H.A. Erlich, N.Arnheim and A.C. Wilson. 1987. Length mutations in human mitochondrial DNA: direct sequencing of enzymatically amplified DNA. Nucl. Acids Res. 15:529-542.