J Forensic Sci, Mar. 2005, Vol. 50, No. 2
Paper ID JFS2004293
Available online at: www.astm.org

TECHNICAL NOTE

Margaret C. Kline,[1] M.S.; Peter M. Vallone,[1] Ph.D.; Janette W. Redman;[1] David L. Duewer,[2] Ph.D.; Cassandra D. Calloway,[3,4] M.S.; and John M. Butler,[1] Ph.D.

Mitochondrial DNA Typing Screens with Control Region and Coding Region SNPs*

ABSTRACT: Mitochondrial DNA (mtDNA) analysis has found an important niche in forensic DNA typing. It is used with highly degraded samples or low-copy number materials such as might be found from shed hair or bones exposed to severe environmental conditions. The primary advantage of mtDNA is that it is present in high copy number within cells and therefore more likely to be recovered from highly degraded specimens. A major disadvantage to traditional forensic mtDNA analysis is that it is time-consuming and labor-intensive to generate and review the 610 nucleotides of sequence information commonly targeted in hypervariable regions I and II (HVI and HVII) of the control region. In addition, common haplotypes exist in HVI/HVII mtDNA sequences that can reduce the ability to differentiate two unrelated samples. In this report we describe the utility of two newly available screening assays for rapid exclusion of non-matching samples. 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. Processing of the LINEAR ARRAY probe panels "mito strips" was automated on a ProfiBlot workstation. Observable variation in 666 individuals is reported and frequencies of the mitotypes within and between populations are presented. 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. These types of screening assays should enable more rapid evaluation of forensic casework samples such that only samples not excluded would be subjected to further characterization through full HVI/HVII mtDNA sequence analysis.

KEYWORDS: forensic science, DNA typing, mtDNA, SSO probes, mitochondrial DNA coding region, primer extension

For the past decade, DNA sequencing of hypervariable regions I (HVI) and II (HVII) in the control region of mitochondrial DNA (mtDNA) has been useful in forensic casework for the analysis of human remains when nuclear DNA systems fail due to low amounts of DNA or highly degraded specimens (1–3). However, considerable effort and expense are required to develop a full HVI and HVII mtDNA sequence (typically positions 16024–16365 and 73–340 are examined). Hence, several mtDNA screening methods have been developed (4–9). These screening methods permit rapid resolution of non-matching samples allowing a laboratory to focus more attention on full-sequencing of samples that cannot be resolved from one another with the screening method. Alternatively, laboratories without DNA sequencing capabilities or expertise can perform mtDNA typing with a screening method and then send only those samples that are not excluded to a private or public laboratory for full HVI/HVII sequencing.

This report details a number of inter-related technologies associated with the use of the newly available Linear Array Mitochondrial DNA HVI/HVII Region-Sequence Typing Kit (Roche Applied Science, Indianapolis, IN). The utility of a new microchip CE method for quantifying PCR products that are then hybridized to immobilized sequence specific oligonucleotide (SSO) probes is demonstrated. A semi-automated sample processing method for the Linear Arrays was developed and is demonstrated. The utility of this mtDNA typing kit is demonstrated by differentiating several hundred unrelated individuals from U.S. Caucasian, African American, and Hispanic groups. These results are compared to previous work with similar mtDNA SSO typing probes. The presence of multiple signals within a probe region resulting from sequence heteroplasmy is described as is the absence of signal within a probe region resulting from failed hybridization of PCR products due to additional destabilizing polymorphisms. In addition, a new coding region single nucleotide polymorphism (SNP) assay was employed to help subdivide 51 individuals possessing the most common Caucasian mtDNA haplotype (10,11).

[1] Biotechnology Division, National Institute of Standards and Technology, Gaithersburg, MD 20899.
[2] Analytical Chemistry Division, National Institute of Standards and Technology, Gaithersburg, MD 20899.
[3] Department of Human Genetics, Roche Molecular Systems, Alameda, CA 94501.
[4] Comparative Biochemistry Group, University of California, Berkeley, CA.

† Parts of this work were presented at the 2003 International Symposium on Human Identification, Phoenix, Arizona, and at the 2004 Annual Meeting of the American Academy of Forensic Sciences, Dallas, Texas.

‡ National Institute of Justice (NIJ) funded this work in part through an interagency agreement with the NIST Office of Law Enforcement Standards and NIJ grant 95-IJ-CX-0014 to Roche Molecular Systems.

Received 31 July 2004; and in revised form 16 Oct. 2004; accepted 16 Oct. 2004; published 2 Feb. 2005.

Materials and Methods

The steps involved in typing the control region SNPs with the mtDNA Linear Arrays are: DNA extraction/quantification, PCR amplification, PCR product quantification, hybridization of the PCR products to the Linear Arrays, detection with colored precipitate, interpretation of results for each sample, comparison of Linear Array results to other samples, and subdividing into groupings. Those samples possessing the most common Caucasian type were also processed with an additional coding region SNP assay as described below.

DNA Samples

A total of 666 DNA samples that had previously been extracted from whole blood were evaluated in this study. These samples, purchased from Interstate Blood Bank Inc. (Memphis, TN), are assumed to be unrelated and were classified into U.S. Caucasian ($N = 286$), African American ($N = 252$), or Hispanic ($N = 128$) groups based on self-declaration. All samples possess unique STR profiles as determined previously with the AmpFℓSTR Identifier STR kit (Applied Biosystems, Foster City, CA) (12).

Bloods were extracted by a modified "saltout" procedure (13). Portions of the genomic extracts were quantified by UV measurement on a Cary Bio 100 double-beam spectrophotometer (Varian Analytical Instruments, Walnut Creek, CA). Based on the UV/260 quantification, the samples were diluted to a nominal 1 ng/μL (14,15). The concentration of the diluted sample was verified by Pico Green quantitation to be between 0.5 ng/μL and 1.5 ng/μL (16).

PCR Amplification

One μL of each sample extract (≈1 ng/μL) was amplified using the mtDNA HVI/HVII Primer mix and mtDNA PCR Reaction Mix supplied by Roche Molecular Systems (Alameda, CA) as part of a beta-test of the Linear Array Mitochondrial DNA HVI/HVII Region-Sequence Typing kit. This kit is now available from Roche Applied Sciences (Indianapolis, IN) in the same format as described in this work (catalog number 03-527-867-001).

Samples are amplified in a duplex PCR that simultaneously generates PCR products for HVI (positions 15975-16418) and HVII (positions 15–429). The HVI primers amplify an approximately 444 bp PCR product using the following primers:

Forward (F15975-15993B)
5′-biotin-CTCCACCATTAGCACCCAA-3′
Reverse (R16418-16401B)
5′-biotin-ATTTCACGGAGGATGGTG-3′

The HVII primers amplify an approximately 416 bp PCR product using the following primers:

Forward (F15-34B)
5′-biotin-CACCCTATTAACCACTCACG-3′
Reverse (R429-410B)
5′-biotin-CTGTTAAAAGTGCATACCGC-3′

Since 1 ng of genomic DNA was used rather than the "Kit" recommended 5 pg (based on nuclear DNA quantification), the number of amplification cycles was reduced from the recommended protocol. Whereas the standard protocol calls for 34–38 cycles, only 28 cycles were run in order to yield PCR products in the desired concentration range (see below). The PCR parameters used with a GeneAmp 9700 thermal cycler (Applied Biosystems) were: 94°C for 14 min, 28 cycles of {92°C for 15 s, 59°C for 30 s, and 72°C for 30 s}, final extension 72°C for 10 min, and hold at 10°C until the products could be removed from the thermal cycler.

Amplification and typing of 11 SNP sites outside the HVI/HVII Region was performed using the multiplex amplification and allele-specific primer extension assay as previously published (11). These sites include SNPs at positions 3010, 4580, 4793, 5004, 7028, 7202, 10211, 12858, 14470 in the coding region and at positions 477 and 16519 in the Variable Regions (control region outside of HVI/HVII) that help differentiate the most common Caucasian Haplotype (10).

Post-PCR Quantification

Forty of the PCR products were quantified by loading 4 μL on to an agarose (3% mass NuSieve, 1% mass SeaKem GTG, Cambrex Bio Science Rockland, Inc., Rockland, ME) yield gel stained with ethidium bromide and imaged/analyzed with a FMBIO III plus (MiraiBio Inc., Alameda, CA). A yield gel quantification standard was supplied from Roche Molecular Systems as part of the beta-site test, which is the same as the DNA Molecular Weight Marker XIV 100 bp ladder available from Roche Applied Sciences (catalog number 1-721-933). A DNA Low Mass Ladder (catalog number 10068-013, Invitrogen, Carlsbad, CA) was also used for PCR product size estimation and quantification.

All 666 HVI/HVII PCR products were quantified using the Agilent 2100 Bioanalyzer Lab Chip system (Agilent Technologies, Palo Alto, CA) and 1 μL of each PCR product according to the manufacturer's protocols. The Agilent 2100 standard ladder calibrates for base pair (bp) sizing, and quantification for up to 12 samples per run (17). Two DNA fragments, 15 bp and 1500 bp, are added in each tested sample as internal standards to quantify PCR product amounts. Quantitation results were imported to an Excel spreadsheet in order to calculate the appropriate quantity of the PCR product to add to each Linear Array probe panel. Typically addition of ≈50 ng total amplified product to each probe panel was targeted.

Control Region SNP Typing

Manufacturer protocols for the Linear Arrays were followed in processing the HVI/HVII PCR products through hybridization to a nylon membrane containing specific probe sequences (Fig. 1). Allele-specific hybridization of the amplified DNA is detected using an enzymatic conversion of a soluble, colorless substrate to a blue-colored precipitate with the detection chemistry originally described by Saiki et al. (18). Cambridge Reference Sequence (19, 20) positions 16093, 16126, 16129, 16270, 16278, 16304, 16309, 16311, and 16362 from HVI and positions 73, 146, 150, 152, 189, 195, 198, 200, and 247 from HVII are probed in this assay.

Samples were processed both manually (120 samples) and using a SLT ProfiBlot workstation (Tecan US, Research Triangle Park, NC) (remaining 446 samples). Steps in the developed Profiblot protocol are listed in Table 1. All reagents required for processing are pre-loaded onto the Tecan Profiblot. Up to 24 probe panels can be processed simultaneously in a 2 h run. The wash solution is placed in a heated/stirred heat block to maintain the required 55°C temperature for hybridization. After the Linear Array probe panels are added to the tray, the Profiblot SLT instrument dispenses the

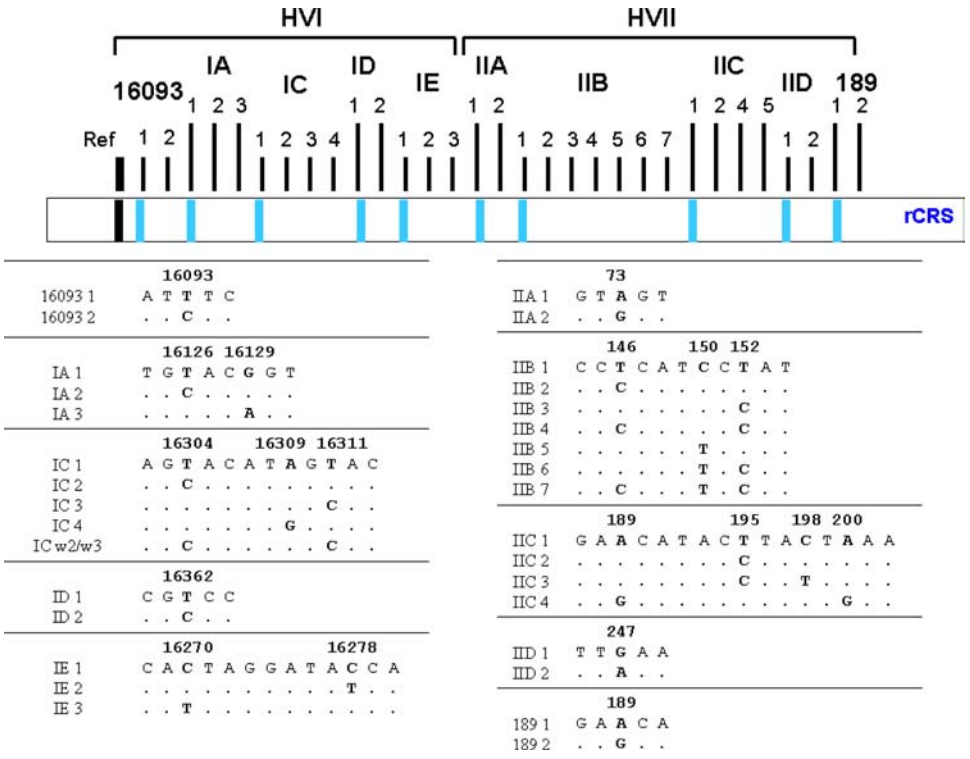

FIG. 1—*Schematic representation of mtDNA LINEAR ARRAY HVI/HVII probe regions and 18 SNPs typed with this assay. Probe 1 for each region corresponds to the revised Cambridge Reference Sequence (rCRS, Andrews et al. 1999). Alternate nucleotides that can be evaluated with each immobilized sequence specific oligonucleotide probe (indicated by their position on the reference sequence) are listed at the bottom of the figure.*

warm wash buffer and pauses for the PCR products to be added to the appropriate well in the 24-position tray. The Linear Arrays were processed through the final water wash without additional human intervention.

Each probe panel was individually aligned with the mtDNA Linear Array Template provided with the kit and the mtDNA types ("mitotypes") for each probe panel were manually assigned following visual inspection of probe signal intensities. For archival purposes, a digital image was taken of the wet strip using a Syngene darkroom (Syngene, Frederick, MD). A yellow light and filter was used to enhance the light blue probe lines on the Linear Array for a black and white image.

Coding Region SNP Typing

Fifty-one samples of the common Caucasian HVI/HVII Linear Array mitotype 1111111111 were typed further with 11 single nucleotide polymorphisms (SNPs) located throughout the mitochondrial genome (mtGenome). This discriminatory SNP panel was developed for resolving haplogroup H1 (315.1C, 263G) (10,21) using a multiplexed SNaPshot assay previously described (11). After amplification of genomic templates, the multiplexed SNaPshot primer extension products were analyzed on an ABI 3100 Genetic Analyzer (Applied Biosystems) using POP-6, a 36-cm array, and an electrokinetic injection of 13 seconds at 1 kV (11). Samples were typed using GeneScan 3.7 and Genotyper 3.7 software programs (Applied Biosystems).

DNA Sequencing

Sequencing of HVI and HVII regions was performed using the primers and amplification protocols described by Wilson et al. (22). Half reactions of the BigDye Version 3 kit (Applied Biosystems) were used following the manufacturer's protocol with approximately 7 ng of ExoSAP (USB Corporation, Cleveland, OH)-treated PCR product. Sequencing products were purified with Performa DTR columns (Edge Biosciences, Gaithersburg, MD) and separated on an ABI 3100 Genetic Analyzer (Applied Biosystems) using POP-6, a 36-cm array, and an electrokinetic injection of 22 seconds at 1 kV. Forward and reverse sequences were aligned and analyzed with Sequencer (GeneCodes, Ann Arbor, MI).

Results and Discussion

A total of 18 polymorphic positions within HVI and HVII are typed with 33 probes striped in 31 lines in the Roche Linear Array Mitochondrial DNA HVI/HVII Region-Sequence Typing kit. The sequence specific oligonucleotide (SSO) probe positions are illustrated in Fig. 1. Probe regions are named following Stoneking et al. (4) and SNP site positions are based on the original mtDNA sequence (19). A reference line ("Ref") is present on the left hand side of each Linear Array to help line it up with the Linear Array Reference Guide provided as part of the kit. Probe 1 in each region represents the nucleotide sequence present in the revised Cambridge Reference Sequence (rCRS) (20). This commercial kit contains more probes than previously published immobilized SSO typing assays (7,8,23–25).

Use of Agilent Chip System for Quantification of PCR Products

The manufacturer PCR protocol specifies 34 amplification cycles with a sample input of (5–10) pg based on a nuclear DNA quantification result. Using these conditions successful amplification and typing of a 2 cm portion of a hair shaft can be accomplished

TABLE 1—*This Tecan Profiblot program was used for performing semi-automated processing of the Roche Linear Array mtDNA HVI/HVII region-sequence typing kit. All reagent solutions are added at the beginning of the two-hour run and samples are fully processed through hybridization, washing, and color development steps. At step 3, a total volume of 30 μL of denatured PCR product and denaturation solution is added (approximately 50 ng total). The same wash buffer is in channel 1 and 2.*

Step	File (Action)	Channel	Temp.	Time	Solution Added
1	Temperature		55°C		
2	Dispense	1	55°C		Wash buffer (2× SSPE, 0.5% SDS, pH 7.4)
3	Pause		55°C		Add denatured PCR product
4	Incubate		55°C	15 min	
5	Aspirate		55°C		
6	Dispense	1	55°C		Wash buffer
7	Aspirate		55°C		
8	Dispense	3	55°C		Streptavidin-Horseradish peroxidase enzyme conjugate
9	Incubate		55°C	5 min	
10	Aspirate		55°C		
11	Dispense	2	55°C		Wash buffer
12	Aspirate		55°C		
13	Dispense	2	55°C		Wash buffer
14	Incubate		55°C	12 min	
15	Aspirate		55°C		
16	Cool		25°C		
17	Dispense	2	25°C		Wash buffer
18	Aspirate		25°C		
19	Dispense	5	25°C		Citrate buffer (0.1 M sodium citrate, pH 5.0)
20	Incubate		25°C	5 min	
21	Aspirate		25°C		
22	Dispense	6	25°C		Color development solution (hydrogen peroxide and TMB)
23	Incubate		25°C	15 min	
24	Aspirate		25°C		
25	Dispense	4	25°C		Deionized water wash
26	Aspirate		25°C		
27	Dispense	4	25°C		Deionized water wash
28	Incubate		25°C	5 min	
29	Aspirate		25°C		
30	Dispense	4	25°C		Deionized water wash
31	End		25°C		

(data not shown). For our population sample testing, we reduced the number of amplification cycles to 28 and used approximately 1 ng of input genomic DNA based on both UV and PicoGreen assays as described in the materials and methods.

The amount of input PCR product is important for reliable performance of the Linear Array assay. If too much PCR product is added to the Linear Arrays, cross-hybridization of the PCR product can occur and make the sample appear like a mixture or heteroplasmic at a particular probe region. If too little PCR product is added, then weak probe signals can make the Linear Arrays hard to interpret or loss of signal may be observed at some of the probe regions. Thus, post-PCR quantification to determine the amount of HVI and HVII PCR products amplified was performed using either an agarose yield gel or a microchip CE system as described in the materials and methods (Fig. 2).

Both the gel based and the chip-based methods appear to be adequate for quantifying the PCR products. Because different controls are used for quantification between gels and chip-based methods, it is important to calibrate the optimal amount of PCR product for the Linear Arrays. The package insert for the Linear Array typing kit recommends addition of approximately (45 to 105) ng of PCR product or 15 μL of PCR product with a concentration of approximately (3 to 7) ng/μL as estimated by comparison with the DNA Molecular Weight Marker XIV run on a yield gel (Fig. 2*A*). In this relatively high DNA concentration range, agarose yield gel measurements have the same accuracy characteristics (bias and precision) as do all other quantitation techniques in common forensic use (26).

In this study we primarily utilized quantification results from the Agilent 2100 BioAnalyzer and its two internal sample controls (Fig. 2*B*). Note that in the Agilent chip-based separation, the HVI and HVII products can be fully resolved from one another and easily quantified. The PCR products displayed in Fig. 2*B* are both approximately 6 ng/μL in concentration. However variability was often observed between the amplification yields for HVI and HVII PCR products (Fig. 3*B*) in which case an average was taken for estimating the amount of DNA that should be added to the Linear Arrays. A sensitivity titration was performed with multiple samples based on the Agilent chip quantification. We observed that approximately 50 ng of PCR product, as quantified by the Agilent system gave optimal results (data not shown). When more than 80 ng of PCR product was added to the Linear Arrays, cross hybridization in the HVIC and HVIIB regions could be seen (data not shown). The Agilent reagent kit used in this study (DNA 1000 assay) specified quantitation accuracy of ±30%; therefore, the Agilent system measured range of (50 to 80) ng represents a true range of (35 to 104) ng.

Figure 3 compares PCR product amounts observed with HVI and HVII PCR products on the 666 tested samples reported here. An average ratio of HVI:HVII amplicons yields of 1.34 was seen in this study with a standard deviation of 0.31. Some representative electropherograms with different HVI/HVII peak ratios are shown in Fig. 3*B*.

A Screen for Samples Possessing the HVI C-stretch

Some samples amplified for the HVI region of the mtDNA control region contain a polymeric stretch of cytosine, usually ranging across nucleotide positions 16184–16193 that is challenging to sequence through (Fig. 2*C*). Explanations of the C-stretch are described as a mtDNA replication problem resulting in length heteroplasmy in the cell and the resulting sequencing products (2,27). Sequencing products typically get out of phase following this C-stretch, leading to poor quality or non-interpretable mtDNA sequencing data. The HVI C-stretch commonly arises when position 16189 contains a cytosine rather than the thymine that is present in the rCRS and a run of 10 or more cytosines occurs (Fig. 2*C*). Detection of the HVI C-stretch is useful if sequencing of that region is required so that different sequencing primers can be used to obtain reliable downstream double-strand sequence reads (28). Not only does the chip-based approach thus permit quantification of HVI and HVII PCR products, this method also enables detection of the presence of the HVI C-stretch when satellite peaks are observed on the tail end of HVI amplicons. These heteroduplex peaks have been previously reported with single-plex HVI PCR products run on a conventional capillary electrophoresis system (29) and can be helpful in selection of a sequencing strategy for HVI samples.

Semi-Automation of Linear Array Sample Processing with Profiblot

The Linear Array mtDNA hybridization assay involves four steps performed sequentially: (1) hybridization of amplified DNA to the LINEAR ARRAY probe panels, (2) binding of

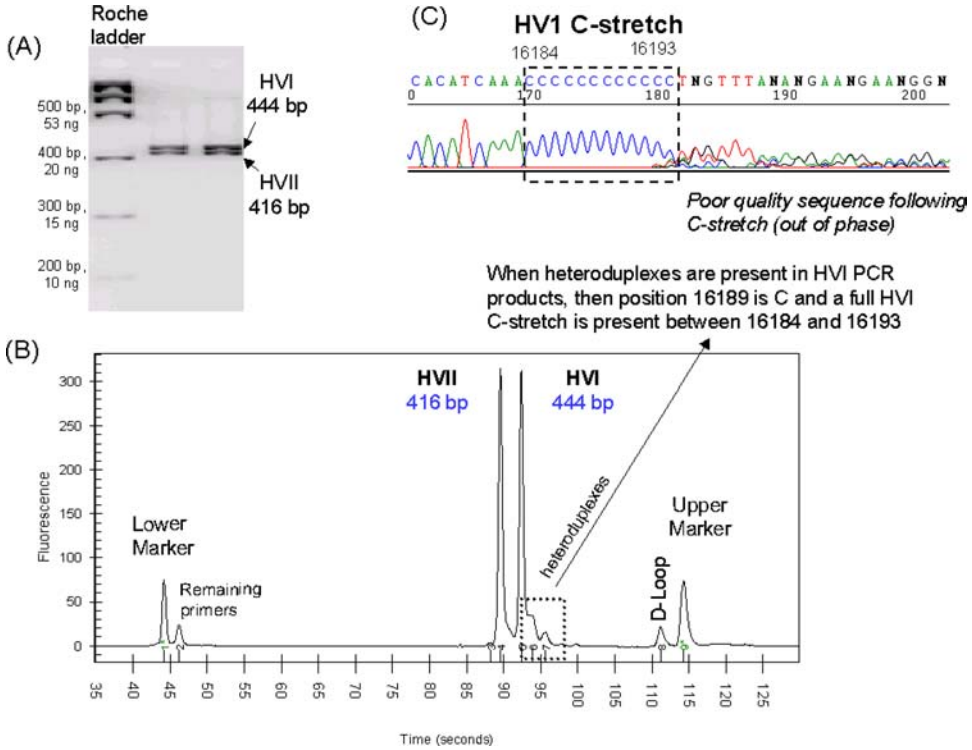

FIG. 2—*(A) Agarose yield gel image of HVI/HVII PCR products compared to Roche Molecular Weight Marker XIV ladder. Quantity of HVI/HVII products was estimated by comparison to the Roche ladder band intensities. (B) Electropherogram from Agilent 2100 Bioanalyzer chip system showing resolution of HVI and HVII products. The lower marker and higher marker are internal standards used for sizing and quantification of PCR products. The presence of heteroduplex peaks following the HVI amplicon are indicative of a stretch of polymeric cytosines that create sequencing problems. (C) Poor sequence quality is observed following the HVI "C-stretch" due to out of phase sequencing products. Different primers can be used on these types of samples to obtain double-strand reads around the C-stretch region.*

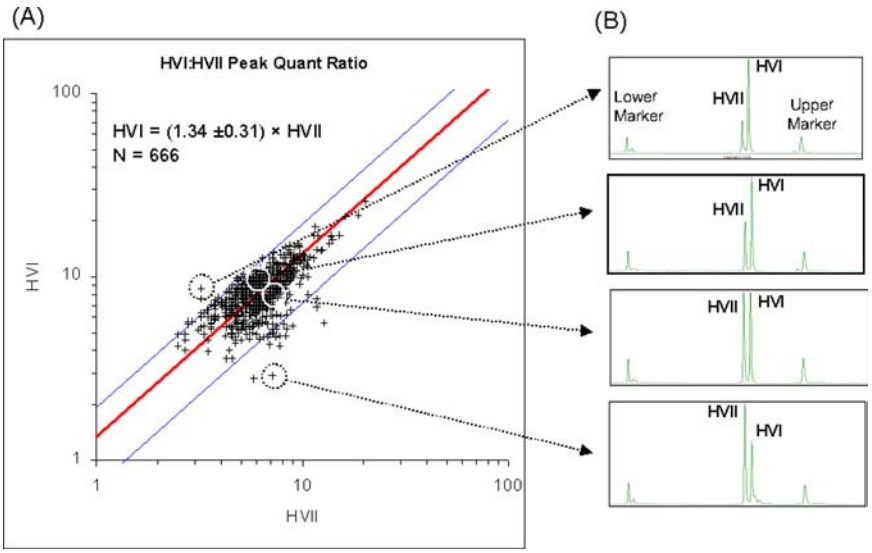

FIG. 3—*(A) Summary of peak quantity ratios for all 666 HVI and HVII PCR products measured on the Agilent Bioanalyzer. (B) Four example electropherograms showing the range of results observed. Most of the samples displayed a higher yield for the HVI PCR products.*

streptavidin-horseradish peroxidase enzyme conjugate to the hybridized biotinylated PCR products, (3) washing the Linear Array probe panels containing immobilized SSO probes to remove non-specifically bound PCR products and enzyme conjugate, and (4) a color development step based on the chemistry described by Saiki et al. (18). Up to 24 samples can be processed as a batch with a 24 well tray. Solutions are typically added manually to each well containing a Linear Array, allowed to incubate in solution for a specific amount of time while the tray is rocked back and forth to incubate the Linear Array in the solution, and then drawn back out of each well in the tray before the next solution is added.

The Tecan Profiblot is an instrument that can dispense solutions in a specified order and maintain those solutions at specified temperatures in order to facilitate processing of SSO probes

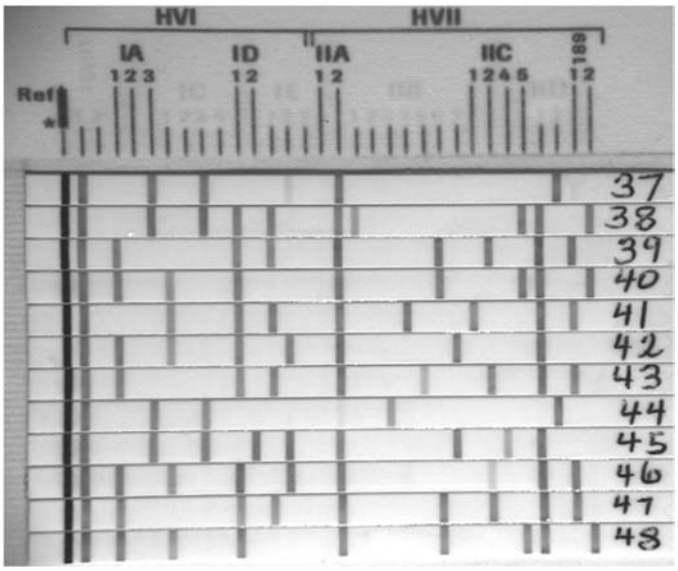

FIG. 4—*Twelve Linear Array mito strips loaded with approximately 50 ng of HVI/HVII PCR product each, quantified by the Agilent system and processed by the Tecan Profiblot workstation.*

immobilized on nylon membranes. The steps of hybridization of the PCR products to the Linear Arrays and subsequent detection with colored precipitate have been automated on a Profiblot workstation using the steps listed in Table 1. Tedious washing steps that are usually performed manually to ensure stringent hybridization of PCR products to the immobilized SSO probes are replaced by the Profiblot protocol. The developed Profiblot program pauses to allow addition of denatured PCR product and then processes the Linear Arrays through the hybridization, addition of enzyme conjugate, washing, and color development steps. Each batch of up to 24 samples takes approximately 2 h to process through these steps. Through automating the fluid changes as well as the temperature changes with various solutions, laboratory personnel can utilize the time their Linear Arrays are being processed to prepare or quantify the next group of samples. Thus (48 to 72) mtDNA samples can be processed to the data interpretation stage in a single 8-hour workday analysis of (2 to 3) runs. Figure 4 shows results from 12 samples processed as part of a batch on the Profiblot.

In contrast, to fully sequence the HVI and HVII regions from the same PCR products would require cleanup of the PCR products (digestion of the remaining primers) followed by additional amplifications with a cycle sequencing reaction kit (four reactions per sample), cleanup of the sequencing reactions, drying the cleaned sequencing products, loading the products on an appropriate sequencer, and analysis of the sequencing data. Thus sequencing analysis of (48 to 72) samples requires collection and analysis of (192 to 288) sequencing files. Using a 16 capillary instrument such as the 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA), sample file collection time alone would be (15 to 22.5) h.

Regardless of whether the immobilized SSO probes are processed manually or using the Profiblot, the hybridization temperature of 55°C and the pH of the hybridization/wash solutions used to process the mtDNA Linear Arrays are critical to obtaining successful results. The pH of the prepared wash solution should be between 7.2 and 7.6. This pH can vary depending upon the pH of the 20X SSPE buffer, and SDS used in solution preparation. If the pH of the wash solution is above 7.6, then probe signal intensities may appear weak and imbalanced.

TABLE 2—*A summary of the frequency and total number of mitotypes observed in this study with 666 individuals and the Roche mtDNA HVI/HVII Linear Array kit. A further breakdown of the common mitotypes by population is found in Table 3. Overall, 185 (27.8%) of the individuals tested had unique mitotypes but 51 (7.7%) of those tested possessed the most common type of 1111111111. A total of 282 different mitotypes were observed in this study. To make the results conservative, "weak" calls are not differentiated from regular probe signals in this study.*

Frequency		Mitotypes	
Times Type Observed	% of Individuals Tested	Number of Types	% of Total mtDNA Types Observed
1	27.8	185	65.6
2	13.8	46	16.3
3	8.1	18	6.4
4	2.4	4	1.4
5	2.3	3	1.1
6	3.6	4	1.4
7	1.1	1	0.4
8	10.8	9	3.2
9	2.7	2	0.7
10	6.0	4	1.4
11	1.7	1	0.4
12	1.8	1	0.4
18	2.7	1	0.4
23	3.5	1	0.4
28	4.2	1	0.4
51	7.7	1	0.4
		282	

Population Typing Results

Table 2 lists the haplotype frequencies observed in the population samples tested. With our 666 U.S. samples, a total of 282 different mitotypes were seen; 185 of these types were observed only once. These unique samples represent 65.6% of the types and 27.8% of the people of our population samples. The most common mitotypes for the three U.S. ethnic groups studied are presented in Table 3. Several mitotypes seem to be found primarily in specific ethnic groups. For example, mitotype 1111111111 is observed 47 times in 286 Caucasians with no occurrences in 252 African Americans and only 4 occurrences in 128 Hispanics.

Table 4 compares this Linear Array data to another SSO typing study performed with North American population samples published by Melton et al. (30). Similar diversity values were observed for each subpopulation even though fewer samples were examined in our study. Despite the addition of 10 more probes and examination of two more regions across HVI and HVII, the most common type, which matches the revised Cambridge Reference sequence, holds at around 7%. This ≈7% figure also exists when complete sequencing is performed across HVI and HVII in Caucasians (10). However, as will be demonstrated below, it is possible to subdivide some of these matching samples with coding region sequence variants.

Observations of "Blanks"

The hybridization of PCR products to immobilized SSO probes can be destabilized by another polymorphism within the probe region. Blanks are reported as a "0" within a mitotype and result from PCR products failing to hybridize to any of the probes

TABLE 3—*Common mitotypes observed in this study broken down by population group. The highest count for each observed mitotype is in bold font. Many of these common types are specific to a particular population in our dataset.*

	Total		Caucasian		African American		Hispanic	
Mitotype	#	% (666)	#	% (286)	#	% (252)	#	% (128)
1111111111	51	7.7	**47**	16.4	0	0	4	3.1
1141224211	28	4.2	0	0	**27**	10.7	1	0.8
1102120111	23	3.5	1	0.3	1	0.4	**21**	16.4
1111121111	18	2.7	**11**	3.8	0	0	7	5.5
1111113111	12	1.8	**11**	3.8	0	0	1	0.8
1111121101	11	1.7	0	0	0	0	**11**	8.6
1121111111	10	1.5	**9**	3.1	1	0.4	0	0
1141220211	10	1.5	0	0	**10**	4.0	0	0
1211121111	10	1.5	**7**	2.4	0	0	3	2.3
1231023220	10	1.5	0	0	**8**	3.2	2	1.6
1111321111	9	1.4	**8**	2.8	1	0.4	0	0
1131121111	9	1.4	**5**	1.7	1	0.4	3	2.3
1101125211	8	1.2	0	0	**7**	2.8	1	0.8
1101125410	8	1.2	0	0	**8**	3.2	0	0
1111325111	8	1.2	**5**	1.7	2	0.8	1	0.8
1211121010	8	1.2	**8**	2.8	0	0	0	0
1211123211	8	1.2	**7**	2.4	0	0	1	0.8
1211126211	8	1.2	**8**	2.8	0	0	0	0
1221121111	8	1.2	**7**	2.4	1	0.4	0	0
1231023020	8	1.2	0	0	**6**	2.4	2	1.6
1330220020	8	1.2	0	0	**7**	2.8	1	0.8
1112111111	7	1.1	**6**	2.1	1	0.4	0	0
1111125111	6	0.9	2	0.7	2	0.8	2	1.6
1111220211	6	0.9	1	0.3	**4**	1.6	1	0.8
1111227410	6	0.9	0	0	**6**	2.4	0	0
1131124111	6	0.9	**4**	1.4	0	0	2	1.6
1131111111	5	0.8	**5**	1.7	0	0	0	0
1311111111	5	0.8	**4**	1.4	0	0	1	0.8
1331121512	5	0.8	0	0	**5**	2.0	0	0
1101126211	4	0.6	0	0	**4**	1.6	0	0
1111112111	4	0.6	**4**	1.4	0	0	0	0
1211121101	4	0.6	2	0.7	0	0	2	1.6
1221120111	4	0.6	**4**	1.4	0	0	0	0

TABLE 4—*Comparison of our dataset with another U.S. population data using similar SSO probes (30). The diversity values are calculated as described in Melton et al. (30). The most common type (MCT) is the same between these data sets and has a similar overall percentage.*

Population	N	# Types	Diversity	Most Common Type	MCT Frequency
Thirty-one probes from 10 regions of HVI/HVII examining 18 SNP sites					
Caucasian	286	116	0.960	1111111111	16.4%
African Am	252	129	0.977	1141224211	10.7%
Hispanic	128	74	0.954	1102120111	16.4%
Total	666	282	0.985	1111111111	7.7%
Twenty-one probes from 8 regions of HVI/HVII examining 13 SNP sites					
Caucasian	922	226	0.964	11111111	15.4%
African Am	805	251	0.983	12112021	6.8%
Hispanic	555	170	0.963	12122011	11.7%
Total	2282	502	0.998	11111111	7.2%

TABLE 5—*Number of "blank" calls by Linear Array probe region observed in this study. Blanks are reported as a "0" within a mitotype; they result from PCR products failing to hybridize to any of the probes due to the presence of additional destabilizing polymorphisms within the probe binding regions. While by convention blanks in the same probe region are considered to represent the same variant, different individuals typing as a blank could have different substitutions (7,30).*

Locus	Number of Blanks Observed	% Observed ($N = 666$)
16093	23	3.5
HVIA	33	5.0
HVIC	76	11.4
HVID	33	5.0
HVIE	60	9.0
HVIIA	3	0.5
HVIIB	96	14.4
HVIIC	122	18.3
HVIID	42	6.3
189	152	22.8

due to the presence of additional polymorphisms not accounted for in the probe possibilities. Different individuals typing as a blank for the same probe region can have different substitutions but for the purposes of data analysis all blanks seen within a particular probe region are considered to represent the same variant (7,30).

The frequencies and probe regions of the observed blanks seen within the 666 population samples examined are listed in Table 5. In our study, we observed that 52% of the samples possessed at least one "0" call and in total 9.6% of all types recorded were blanks. The probe regions with the highest frequency of blanks, HVIC, HVIIB, and HVIIC (Table 5) were noted by Budowle et al. (31) to be most likely to have closely spaced polymorphisms in U.S. Caucasians, African Americans, and Hispanics.

TABLE 6—*Possible heteroplasmic sites detectable from HVI and HVII with potential probe results using the Linear Array kit and a summary of the 7 heteroplasmic sites observed in this study. Different probe combinations are possible in some instances due to the presence of other polymorphic bases present near the heteroplasmic site. The observed instances of heteroplasmy have been confirmed with DNA sequencing.*

Site	Nucleotides (rCRS/variant)	Potential Heteroplasmies	Number Observed in 666 U.S. Samples
16093	T/C	16093 1&2	1
16126	T/C	IA 1&2	
16129	G/A	IA 1&3	
16270	C/T	IE 1&3	
16278	C/T	IE 1&2	
16304	T/C	IC 1&2	
16309	A/G	IC 1&4	
16311	T/C	IC 1&3	
16362	T/C	ID 1&2	1
73	A/G	IIA 1&2	
146	T/C	IIB 1&2 or IIB 3&4	2 (both IIB 3&4)
150	T/C	IIB 1&5	
152	T/C	IIB 1&3 or IIB 5&6 or IIB 2&4	2 (1 IIB 2&4, 1 IIB 5&6)
189	A/G	189 1&2	1
195	T/C	IIC 1&2	
198	C/T	IIC 1&3	
247	G/A	IID 1&2	

Heteroplasmy Detected with Linear Arrays

Sequence heteroplasmy, or the presence of two nucleotides at a single position, can be detected with the Linear Array assay at 17 locations in HVI and HVII. Table 6 lists the probe combinations necessary to detect possible heteroplasmic sites in HVI and HVII PCR products. In the 666 samples examined here, we observed 7 instances of heteroplasmy at 5 different sites: 16093, 16362, 146, 152, and 189 (Table 6). These sites are among the fast evolving homoplasmic sites (21,32) and reported heteroplasmic "hotspots" (5,24), which is why many of these probe variants were selected for the Linear Array assay.

Multiple probe combinations are possible in some instances due to the presence of other polymorphic bases near the heteroplasmic site within a probe region. For example, position 152, which is present in probe region HVIIB, can be detected as a C/T heteroplasmy if probes HVIIB 1 and 3 are detected in the same sample. In this scenario, position 146 is a T and position 150 is a C. However, if probes HVIIB 2 and 4 are observed, then position 152 is heteroplasmic but position 146 is a C, and position 150 is a T. Alternatively, if probes HVIIB 5 and 6 are observed, then position 152 is heteroplasmic, position 146 is a T and position 150 is a T. DNA sequencing has been performed on the heteroplasmic samples seen in this study to verify that both nucleotides are present at the site in question (data not shown).

Subdividing the Most Common Type with Coding Region SNPs

As noted above (Tables 2–4), approximately 7% of the samples possess a common mtDNA haplotype with the Linear Array assay that cannot be resolved by analyzing only the HVI and HVII SSO probe regions. Full sequencing of HVI and HVII may be able to subdivide a few more of these samples but regardless many of these samples will likely remain undistinguishable from one another based on HVI and HVII sequences alone (10,21). Expansion of mtDNA sequencing and typing to the full mtDNA genome permits use of polymorphisms that occur outside the control region. These coding region SNPs can provide limited resolution to separate many individual samples that cannot be subdivided with control region information alone.

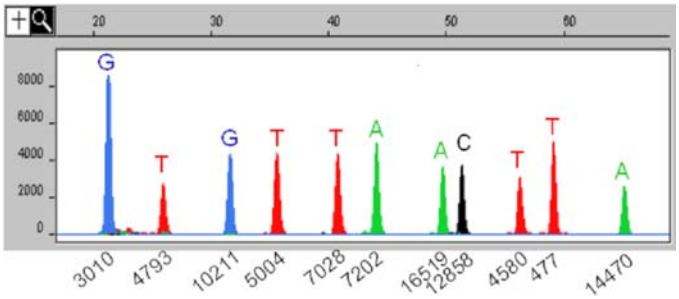

FIG. 5—*Electropherogram of the 11 plex allele-specific primer extension assay for subdividing individuals containing the most common Caucasian mtDNA type. In this study using 666 samples, the Linear Array most common type of 1111111111 was observed in 51 individuals (7.7%).*

TABLE 7—*Breakdown of the results from an 11-plex allele-specific primer extension assay used to subdivide the 51 individuals observed with the most common Caucasian type (1111111111 with the Roche Linear Array). A total of 12 different subtypes were observed, five of them existing in only one sample. The left-most column shows the nucleotide positions examined around the mtGenome preceded by the base present in the revised Cambridge Reference Sequence (rCRS). The polymorphic bases, relative to the rCRS nucleotide, are in bold font.*

	Number of Times Observed											
	1	1	1	1	1	2	2	3	4	4	15	16
G3010	***A***	G	G	G	G	***A***	G	G	G	G	***A***	G
A4793	A	A	A	A	A	A	A	A	A	***G***	A	A
C10211	C	C	C	C	C	C	C	C	C	C	C	C
T5004	T	***C***	T	T	T	T	T	T	T	T	T	T
C7028	C	C	C	**T**	**T**	C	***T***	C	***T***	C	C	C
A7202	A	A	A	A	A	A	A	A	A	A	A	A
T16519	***C***	T	**C**	**C**	T	**C**	**C**	T	T	**C**	**C**	**C**
C12858	***T***	C	C	C	C	C	C	C	C	C	C	C
G4580	G	G	G	G	G	G	***A***	G	***A***	G	G	G
T477	***C***	T	T	T	T	***C***	T	T	T	T	T	T
T14470	T	T	***A***	T	T	T	T	T	T	T	T	T

The most common mitotype 1111111111 was further subdivided into 12 different types using an 11 plex mtDNA coding region SNP assay. This assay uses multiplex PCR and a multiplex allele-specific primer extension assay to simultaneously examine 11 widely spaced polymorphic nucleotides around the mtDNA genome (Fig. 5). These sites were selected to help further resolve the most common Caucasian HVI/HVII as determined by full mtGenome sequencing experiments (10,11).

Table 7 contains a summary of the results observed from 51 individuals, representing 47 Caucasian and 4 Hispanic samples, possessing the most common type. Note that two of the sites 10211 and 7202 did not exhibit any variation in the 51 samples. However with variation observed in the other 9 tested SNPs, 12 different types were observed with 5 being unique to an individual sample and an additional 5 containing 4 or fewer individuals in them. Unfortunately, two larger groups containing 15 and 16 individuals could not be subdivided with this 11 plex screening assay.

Conclusions

We have demonstrated that mtDNA screening assays from the control region and coding region can be used to effectively differentiate unrelated individuals. A protocol has been developed for

processing Linear Array mtDNA HVI/HVII probe panels in a semi-automated fashion. Furthermore, the use of coding region SNPs can help subdivide individuals possessing the most common type seen in Caucasian individuals.

Acknowledgments

The authors thank Lois Tully for encouragement and support of this project and Mike Coble for his helpful comments on the manuscript. The coding region SNP assay resulted from collaboration between our group at NIST and the Armed Forced DNA Identification Laboratory; the work of Mike Coble, Rebecca Just, and Tom Parsons is gratefully acknowledged in identifying these informative SNP sites.

References

1. Wilson MR, Stoneking M, Holland MM, DiZinno JA, Budowle B. Guidelines for the use of mitochondrial DNA sequencing in forensic science. Crime Lab Digest 1993;20(4):68–77.
2. Holland MM, Parsons TJ. Mitochondrial DNA sequence analysis—validation and use for forensic casework. Forensic Sci Rev 1999; 11(1):22–50.
3. Budowle B, Allard MW, Wilson MR, Chakraborty R. Forensics and mitochondrial DNA: applications, debates, and foundations. Annu Rev [PubMed] Genomics Hum Genet 2003;4:119–41.
4. Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA. Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. [PubMed] Am J Hum Genet 1991;48:370–82.
5. Tully G, Sullivan KM, Nixon P, Stones RE, Gill P. Rapid detection of mitochondrial sequence polymorphisms using multiplex solid-phase [PubMed] fluorescent minisequencing. Genomics 1996;34:107–13.
6. Butler JM, Levin BC. Forensic applications of mitochondrial DNA. Trends Biotech 1998;16:158–62.
7. Reynolds R, Walker K, Varlaro J, Allen M, Clark E, Alavaren M, et al. 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 2000;45(6):1210– [PubMed] 31.
8. Gabriel MN, Calloway CD, Reynolds RL, Andelinovic S, Primorac D. 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 [PubMed] 2001;42(3):328–35.
9. Budowle B, Planz JV, Campbell RS, Eisenberg AJ. Single nucleotide polymorphisms and microarray technology in forensic genetics—development and application to mitochondrial DNA. Forensic Sci Rev 2004;16:21–36.
10. Coble MD, Just RS, O'Callaghan JE, Letmanyi IH, Peterson CT, Irwin JA, et al. Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians. Int J [PubMed] Legal Med 2004;118:137–46.
11. Vallone PM, Just RS, Coble MD, Butler JM, Parsons TJ. A multiplex allele specific primer extension assay for 11 forensically informative SNPs distributed throughout the mitochondrial genome. Int J Legal Med [PubMed] 2004;118:147–57.
12. Butler JM, Schoske R, Vallone, PM, Redman JW, Kline MC. Allele frequencies for 15 autosomal STR loci on U.S. Caucasian, African American, and Hispanic populations. J Forensic Sci 2003;48(4):908–11. [PubMed]
13. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res [PubMed] 1988;16(3):1215.
14. Beaven GH, Holiday ER, Johnson EA. Optical properties of nucleic acids and their components. In: Chargaff E, Davidson JN, editors. The nucleic acids. New York: Academic Press, 1955;Volume I:493–553.
15. Wilfinger WW, Mackey K, Chomczynski P. Effect of pH and ionic strength on the sprctrophotometric assessment of nucleic acid purity. BioTechniques 1997;22:474–81. [PubMed]
16. Singer VL, Jones LJ, Yue ST, Haugland RP. Characterization of PicoGreen reagent and development of a fluorescence-based solution assay for double-stranded DNA quantitation. Anal Biochem 1997;249(2):228–38. [PubMed]
17. Panaro NJ, Yuen PK, Sakazume T, Fortina P, Kricka LJ, Wilding P. Evaluation of DNA fragment sizing and quantification by the Agilent 2100 Bioanalyzer. Clinical Chem 2000;46(11):1851–3.
18. Saiki RK, Walsh PS, Levenson CH, Erlich HA. Genetic analysis of amplified DNA with immobilized sequence- specific oligonucleotide probes. Proc Natl Acad Sci USA 1989;86:6230–4. [PubMed]
19. Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. Nature 1981;290:457–65. [PubMed]
20. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 1999;23(2):147. [PubMed]
21. Allard MW, Miller K, Wilson M, Monson K, Budowle B. Characterization of the Caucasian haplogroups present in the SWGDAM forensic mtDNA dataset for 1771 human control region sequences. J Forensic Sci 2002;47:1215–23. [PubMed]
22. Wilson MR, Polanskey D, Butler JM, DiZinno JA, Replogle J, Budowle B. Extraction, PCR amplification, and sequencing of mitochondrial DNA from human hair shafts. BioTechniques 1995;18:662–9. [PubMed]
23. Comas D, Reynolds R, Sajantila A. Analysis of mtDNA HVRII in several human populations using an immobilised SSO probe hybridisation assay. Eur J Hum Genet 1999;7(4):459–68. [PubMed]
24. Calloway CD, Reynolds RL, Herrin GL, Jr., Anderson WW. The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. Am J Hum Genet 2000;66(4):1384–97. [PubMed]
25. Gabriel MN, Calloway CD, Reynolds RL, Primorac D. Identification of human remains by immobilized sequence-specific oligonucleotide probe analysis of mtDNA hypervariable regions I and II. Croat Med J 2003;44(3):293–8. [PubMed]
26. Kline MC, Duewer DL, Redman JW, Butler JM. NIST Mixed Stain Study #3: DNA quantification practice and it's influence on short tandem repeat multiplex performance. Anal Chem 2003;75(10):2463–9. [PubMed]
27. Parson W, Parsons TJ, Scheitnauer R, Holland MM. Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case. Int J Legal Med 1999;112(5):348–9. [PubMed]
28. Rasmussen EM, Sorensen E, Eriksen B, Larsen HJ, Morling N. Sequencing strategy of mitochondrial HV1 and HV2 DNA with length heteroplasmy. Forensic Sci Int 2002;129(3):209–13. [PubMed]
29. Butler JM, Wilson MR, Reeder DJ. Rapid mitochondrial DNA typing using restriction enzyme digestion of polymerase chain reaction amplicons followed by capillary electrophoresis separation with laser-induced fluorescence detection. Electrophoresis 1998;19(1):119–24. [PubMed]
30. Melton T, Clifford S, Kayser M, Nasidze I, Batzer M, Stoneking M. Diversity and heterogeneity in mitochondrial DNA of North American populations. J Forensic Sci 2001;46(1):46–52. [PubMed]
31. Budowle B, Wilson MR, DiZinno JA, Stauffer C, Fasano MA, Holland MM, Monson, KL. Mitochondrial DNA regions HVI and HVII population data. Forensic Sci Int 1999;103(1):23–35. [PubMed]
32. Stoneking M, Hypervariable sites in the mtDNA control region are mutational hotspots. Am J Hum Genet 2000;67(4):1029–32. [PubMed]

Additional information and reprint requests:
Margaret Kline, M.S.
Research Biologist
National Institute of Standards and Technology
100 Bureau Dr., Mail Stop 8311
Gaithersburg, MD 20899-8311
E-mail: margaret.kline@nist.gov