Rebecca Reynolds,[1] Ph.D.; Karen Walker,[1] B.A.; Joseph Varlaro,[1,2] B.S.; Marie Allen,[1,3] Ph.D.; Ellen Clark,[1,4] Ph.D.; May Alavaren,[1,5] B.A.; and Henry Erlich,[1] Ph.D.

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

REFERENCE: Reynolds R, Walker K, Varlaro J, Allen M, Clark E, Alavaren M, Erlich H. 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):1210–1231.

ABSTRACT: We have developed a rapid, immobilized probe-based assay for the detection of sequence variation in the hypervariable segment II (HVII) of the mitochondrial DNA (mtDNA) control region. Using a panel of 17 sequence-specific oligonucleotide (SSO) probes immobilized on nylon membrane strips, we typed 689 individuals from four population groups. The genetic diversity value for each population was calculated from the frequency data, and the frequencies of distinct "mitotypes" in each group were determined. We performed DNA sequence analysis of 129 samples to characterize the sequences associated with "blanks" (absence of probe signals) and weak probe signals. Out of 689 samples, we observed five heteroplasmic samples (excluding the variable C-stretch beginning at position 303) using the immobilized SSO probe panel. The SSO probe strips were used for the analysis of shed hairs and bloodstains from several criminal cases in Sweden, one of which is described here. We conclude that this mtDNA typing system is useful for human identification and significantly decreases casework turnaround time.

KEYWORDS: forensic science, mitochondrial DNA, HVII, immobilized sequence-specific oligonucleotide probes, DNA typing—human

Mitochondrial DNA sequence variation among individuals within a population and between populations has been studied extensively by direct DNA sequence analysis and restriction enzyme analysis (1–11) and by using a panel of sequence-specific oligonucleotide (SSO) probes following PCR amplification (12–18). Some of these studies revealed that the control region of the mitochondrial genome is highly polymorphic and that the majority of the sequence variation is localized in two ~450-base pair segments (2,4,5). These segments are referred to as hypervariable regions I and II (HVI and HVII). In addition to its high degree of polymorphism, the mitochondrial genome has other features that make it particularly suitable for the analysis of samples that are not amenable to nuclear marker typing (e.g., shed hairs, mass disaster and burial remains, and missing persons). First, individual cells can contain hundreds or thousands of copies of the mitochondrial genome because each mitochondrion carries multiple copies of its genome and each cell contains many mitochondria (19). Consequently, mtDNA persists longer than nuclear DNA as biological samples age and degrade, and so frequently it is possible to PCR amplify and type mtDNA from samples that cannot be typed using nuclear marker tests (e.g., HLA DQA1, D1S80, STRs). Another unique feature of mtDNA is that it is inherited matrilineally (20). This feature most dramatically impacts missing person and mass disaster cases because the missing or unidentified person's biological mother, siblings, and maternal relatives will all have the same mtDNA sequence, with few exceptions. Therefore, these individuals can provide reference samples for the missing or unidentified individual.

Most of the laboratories currently engaged in mtDNA research and typing use direct DNA sequence analysis of PCR products amplified from various regions of the mitochondrial genome. Stoneking et al. (12) developed an alternative method for screening large numbers of samples using sequence-specific oligonucleotide (SSO) probes. Their dot blot test consists of 23 SSO probes spanning nine regions within the two hypervariable segments of the mtDNA control region. The authors reported that the degree of diversity revealed by this panel of probes is only slightly less than the diversity revealed by direct DNA sequence analysis. While the direct DNA sequence analysis and dot blot approaches to mtDNA typing provide a significant amount of valuable information, the procedures and data analysis are very time-consuming. Samples must be sequenced in both directions and, frequently, sequenced a second time to resolve ambiguities. With Stoneking et al.'s dot blot approach, 23 individual hybridization reactions must be performed and analyzed. Consequently, neither of these approaches is practical for routine forensic casework.

To address the need for a rapid, easy-to-interpret method for mtDNA typing, we have developed an immobilized SSO probe-based assay. This assay has essentially the same format as the

[1] Human Genetics Department, Roche Molecular Systems, Inc., Alameda, CA.
[2] Current address: Boston Police, Crime Laboratory Unit, 1 Schroeder Plaza, Boston, MA.
[3] Department of Genetics and Pathology, University of Uppsala, Sweden.
[4] Current address: California Department of Justice, DNA Laboratory, 626 Bancroft Way, Berkeley, CA.
[5] Current address: Mosaic Technologies, Inc., 1106 Commonwealth Ave., Boston, MA. Genomics Collaborative, 99 Erie St., Cambridge, MA.

* This work was supported in part by NIJ Grant 95-IJ-CX-0014 to RR and HE.
Received 1 April 1999; and in revised form 3 Dec. 1999; accepted 24 Jan. 2000.

widely used AmpliType® PM and HLA DQA1 forensic test kits. This paper describes the results obtained from typing 689 unrelated individuals using a panel of 17 immobilized sequence-specific oligonucleotide probes designed to detect sequence polymorphisms in the HVII region of the mitochondrial genome. The probes are complementary to sequence variants within the five regions of HVII described by Stoneking et al. (12). Within each probe binding region, one of four categories of probe signal was observed: (1) a single probe is positive, (2) a single probe signal is visible but its intensity is weaker than a positive signal, (3) no probe signals are visible, or (4) two probe signals are visible. These categories were characterized by direct DNA sequence analysis of over 120 samples, and the results were used to prepare guidelines for the interpretation of mitotypes obtained with the immobilized SSO probe system. We also describe the results obtained from shed hairs collected from a homemade bomb and demonstrate the utility of this assay for forensic casework. A statistical description of population variation and a list of observed mitotypes were generated for each of the four population groups from which samples were collected. The genetic diversity values range from 0.95 to 0.98 across these populations and correspond to the level of discrimination provided by the HLA DQA1 and D1S80 assays.

Another valuable application of the immobilized SSO probe-typing system that will be described elsewhere is the analysis of remains recovered from mass graves. The use of this mtDNA typing system for the identification of human remains will depend on the availability of mitotype diversity data for the relevant populations (see above). Obtaining such population mtDNA diversity data can be facilitated by the use of this simple and rapid typing assay.

Materials and Methods

Population Database Samples

The 200 U.S. Caucasian, 200 African American, and 200 U.S. Hispanic samples used in this study were generously provided by Dr. Marcia Eisenberg (Roche Biomedical Laboratories, now Laboratory Corporation of America). The 89 Japanese samples were provided by Takehiko Sasazuki (Kyushu University, Fukuoka, Japan). All of the DNA samples used in this study were extracted from blood.

Amplification Conditions

The quantity of mtDNA in the database samples was not known; there is no simple method for determining mtDNA quantity. A volume containing ~1 ng of nuclear DNA from each sample was added to the PCR amplification reaction. Nuclear DNA was quantitated using the QuantiBlot™ kit (PE Biosystems).

Mitochondrial DNA was amplified in reactions containing 12 m*M* Tris-HCl, pH 8.3, 60 m*M* KCl, 2.4 m*M* $MgCl_2$, 0.2 m*M* each dNTP, 250 U/mL AmpliTaq Gold® DNA Polymerase (PE Biosystems), and 0.2 μM each primer [5'-XCACCCTATTAACCACTCACG (HVII-L) and 5'-XCTGTTAAAAGTGCATACCGCCA (HVII-R) where X= biotin]. To activate the AmpliTaq Gold DNA Polymerase, the reactions were heated at 92°C for 12 min prior to cycling. Following enzyme activation, the samples are amplified for 34 cycles in the DNA Thermal Cycler 480 (PE Biosystems) at 92°C for 1 min, 60°C for 30 s, and 72°C for 30 s with a final extension step at 72°C for 7 min. Samples can be amplified in the GeneAmp PCR System 9600 (PE Biosystems) using the same parameters except the denaturation step is 30 s instead of 1 min and the final extension step is 10 min instead of 7 min.

The quantity of the amplified 415 bp HVII PCR product can be evaluated on a 2% agarose gel in 1X TBE containing 0.5 μg/mL ethidium bromide in both gel and running buffer. A 10-cm gel is run at 115 to 120 V for 20 to 30 min.

HVII Probe Sequences

The sequences for the 17 SSO probes immobilized on nylon membrane strips are listed in Table 1. The asterisks indicate probes that are complementary to the sense strand. The positions spanned by each probe are listed in the right column, following the numbering scheme of the reference mtDNA sequence (21). The sequences that are distinguished by the probes in regions A-E are listed in Table 2.

The choice of sequences to be probed was based on our goal to maximize the genetic diversity values using as few probes as possible. We queried mtDNA sequence databases provided by the FBI, AFDIL, and C. Ginther with probe sequences described by Stoneking et al. (12) for both HVI and HVII to estimate the contribution of each probe to the overall genetic diversity values. We found that HVI has a higher level of sequence diversity relative to HVII. However, it takes fewer probes in the HVII region to reach a given level of genetic diversity because the variable positions within HVII are more clustered than the variable positions within HVI. To further increase the genetic diversity value obtained with this assay, we designed additional probes for HVII using unpublished mtDNA sequence and frequency information that was generously provided by Rhonda Roby and Charity Diefenbach. We are very grateful for this information, which led to the design of probes B4–B7, C4, and C5.

Generally, for a particular sequence, probes are designed to be complementary to the strand that creates the most destabilizing mismatch with the alternative sequences that define the other variants or alleles. The more destabilizing the mismatch, the lower the potential for cross-hybridization. The probes for region A, which distinguish A and G at position 73 of the mtDNA reference sequence (21), provide an example of how the nature of the mismatch influences probe design. The A1 probe listed in Table 1 is complementary to the anti-sense strand and forms an A-T match at position 73 with HVII PCR product amplified from a sample containing the A1 sequence. The A2 sequence has a G residue at position 73 and forms an A-C mismatch with the A1 probe. If the A1 probe had been designed to be complementary to the sense strand, then the T residue in the probe would form a G-T mismatch with the A2 sequence. A G-T mismatch is much less destabilizing than an A-C mismatch (22) and can lead to detectable cross hybridization. In contrast, the A2 probe is complementary to the opposite (sense) strand. This sequence was chosen because it creates an A-C mismatch with the A1 sequence. If the A2 probe had been designed to be complementary to the anti-sense strand, a less destabilizing G-T mismatch would be formed with the A1 PCR product.

While this approach generally works, sometimes modifications to the initial probe sequence are necessary to achieve strong, specific signals. For example, sometimes the probe is redesigned to be complementary to the opposite strand because, for reasons that are not apparent, the less destabilizing mismatch provides a better balance between our criteria of strong signal and weak/no cross-hybridization. Occasionally, it is necessary to substitute a base in the probe sequence to create a mismatch with the intended target sequence. Three of the probe sequences in Table 1 are not perfectly complementary to the corresponding HVII sequence (note bases en-

TABLE 1—*HVII probe sequences.*

PROBE	10	20	30	POSITION
A1	5' G G G G T A T G C A C G C G A T A G			68-85
A2	5' C T A T C G C G T G C A C A C C C C *			68-85
B1	5' T A A A T A A T A G G A T G A G G C A G G *			140-160
B2	5' T T C C T G C [A] C C A T C C T A T T A T T T A T			138-161
B3	5' C T G C C T C A T C C C A T T A T T T A			141-160
B4	5' C T G C C C C A T C C C A T T A T T			141-158
B5	5' A A A T A A T A G A A T G A G G C A G G A A *			138-159
B6	5' T A A A T A A T G G A A T G A G G C A G *			139-160
B7	5' A A A T A A T G G A A T G G G G C A G G *			140-159
C1	5' C A C T T T A G T A A G T A T G T T C G C *			185-205
C2	5' C G A A C A T A C C T A C T A A A G T G T [T]			186-207
C4	5' A C T T T A A T A G G T A T G T T C G C C *			184-204
C5	5' C T T C A G T A A G T A T G C T C G C *			185-203
D1	5' T G T G C A G A C A T T C A A T T G T T A T T A *			236-259
D2	5' A A C A A T T A A A T G T C T G C A C A G			240-260
E2	5' G [T] G G A G **Z** G G G G G G T T T G G *			298-314(+1C)
S	5' C A A A G A A C C C T A A C A C C A G C C T A A C C A G A T T T C			362-394

Probe sequences are written 5' to 3'. Probes designated with an asterisk are complementary to the sense strand. Bases enclosed in boxes are not complementary to the target sequence. These base substitutions are necessary for achieving a strong signal with little or no hybridization with other target sequences. See Materials and Methods for further discussion.

TABLE 2—*Sequence variation in HVII distinguished by 17 immobilized SSO probes.*

PROBE DESIGNATION*	SEQUENCE VARIATION DETECTED
	73
A1	G T A T G
A2	- - G - -
	146 150 152
B1	C C T C A T C C T A T
B2	- - C - - - - - - - -
B3	- - - - - - - - C - -
B4	- - C - - - - - C - -
B5	- - - - - - T - - - -
B6	- - - - - - T - C - -
B7	- - C - - - T - C - -
	189 195 198 200
C1	G A A C A T A C T T A C T A A A
C2	- - - - - - - - C - - - - - - -
C4	- - - - - - - - C - - T - - - -
C5	- - G - - - - - - - - - - G - -
	247
D1	T T G A A
D2	- - A - -
	309 310
E2	A A C C C C C C C C T C C A C

* These probe designations were chosen to be consistent with the designations assigned to particular sequence variants by Stoneking, et al. (12). All probes designated "1" are specific for the reference mtDNA sequence (18). In regions A and D, our two probes in each of these regions distinguish the same sequence variants as their two probes. In region B, their probe panel distinguishes 3 sequence variants. Our probe panel also distinguishes these 3 variants (probes B1-B3) as well as an additional four sequence variants we have designated B4-B7. In region C, their probe panel distinguishes 3 sequence variants. We designed probes for two of these sequence variants (C1 and C2) and added two probes to distinguish two additional sequence variants (C4 and C5). Consequently, the C3 sequence variant will not hybridize to our probe panel and their probe panel will not detect the C4 and C5 variants. As with the probes in regions A-D, the single E region probe is numbered to be consistent with the sequence variants defined by Stoneking, et al. (12) and carries the designation "2" instead of "1" because it is specific for eight C residues rather than seven.

closed in boxes). Less frequently, a modified nucleotide needs to be incorporated into the probe to produce higher signal intensity without increasing the level of cross-hybridization. For example, the E2 probe contains a deaza-dG residue (designated "**Z**" in the table).

These examples of deviation from complementarity demonstrate that probe design is highly empirical. While T_m values may be used as a guideline for initial probe design, frequently there is little correlation between the T_m values calculated from base content (i.e., overall numbers of each base) and the performance of a probe. This finding is not surprising given that neither the primary nor the secondary structure of the probe is taken into account by these T_m calculations. In addition, we have observed that a probe attached to a nylon membrane via its 3' end does not necessarily perform the same as when it is attached via its 5' end. These results suggest that immobilized probes are somehow constrained and therefore cannot form all possible conformations. This aspect of immobilized probes may also contribute to the frequent discrepancy between predicted performance based on T_m calculations that assume DNA is free in solution and actual performance in the hybridization assay.

HVII Typing Protocol

The HVII typing protocol is the same as the AmpliType PM/PM+DQA1 typing protocol (available from PE Biosystems) with the following modifications:

1. PCR product is prepared for hybridization by mixing 20 μL of HVII PCR product with 20 μL base denaturation solution (16 mg/mL NaOH pellets, 30 mg/mL EDTA, 50 μg/mL thymol blue) rather than by heat denaturation in the presence of EDTA.
2. The assay is performed at 54°C ± 1°C rather than at 55°C ± 1°C.
3. The Wash Solution contains 2X SSPE, 0.5% SDS rather than 2.5X SSPE, 0.1% SDS.
4. Less SA-HRP conjugate per strip is used: 8 μL instead of 27 μL.
5. The volume of all typing solutions used is reduced to 3 mL (smaller strips and tray wells).

HVII typing can also be performed using the Profiblot IIT automated strip developer (SLT).

DNA Sequence Analysis

Some of the DNA samples were sequenced using the HVII primers described above and the AmpliCycle® DNA Sequencing Kit (PE Biosystems) with [^{33}P] dATP. Samples were loaded on a 35-cm 6% acrylamide/7 *M* urea gel and run in 1X TBE for 1 h 40 min to 3 h when the HVII-L primer was used. Gels were dried and exposed to Biomax film (Kodak). Many of these samples were not sequenced beyond the C-stretch that begins at position 303 because the goal was to confirm the sequence variation in regions A-D. Samples also were sequenced with the HVII primers using the Big Dye sequencing kit (PE Biosystems) and the ABI Prism® 310 Genetic Analyzer according to manufacturer's protocols.

Casework Analysis

Evidentiary and reference hairs were washed with water prior to extraction. DNA was extracted by cutting individual hairs and incubating the fragments in 200 μL PCR buffer containing DTT (50 mM KCl, 10 mM Tris-HCl, pH 8.3, 1.5 m*M* MgCl2, 0.45% Nonidet P40, 0.45% Tween 20, and 35 m*M* DTT). Five μL proteinase K (10 mg/mL) was added and the sample was incubated at 56°C for 2 h. The proteinase K was inactivated at 95°C for 10 min (23).

All extracted samples were amplified and typed using the immobilized SSO probe strips at two different times in duplicate using the procedures described above. For DNA sequence analysis, samples were amplified using the HVI primer pair L15997 and M13(-21)-H16401 and the HVII primer pair L29 and M13(-21)-H408. Primer sequences and PCR conditions previously described (24) were used with the exception that the PCR was performed without nesting for 40 cycles. These reactions were carried out in duplicate at two different times. The TaqFS Dye Primer Sequencing Kit (PE Biosystems) with the M13(-21) tailed primers and the TaqFS Dye Terminator Sequencing Kit with the upstream primers were used for sequencing in both directions. Sequencing reactions were analyzed on an ABI 373 DNA sequencer and the SeqEd package was used for data analysis.

Results

PCR Amplification and DNA Probe Strip Typing of HVII Sequence Variants

The mtDNA typing assay consists of two steps that can be completed in 4 to 5 h. First, a 415-base pair (bp) region of the HVII

segment of the mtDNA control region is amplified using a pair of biotinylated primers. The second step involves hybridization of the PCR products to a panel of immobilized sequence-specific oligonucleotide (SSO) probes followed by a color development reaction. Probes that have hybridized to specific PCR products are visualized as a blue line on the strip; the pattern of probe signals on a strip constitutes an individual's "mitotype." For samples with severely degraded DNA, the region can be amplified as two smaller pieces to increase the amplification success rate. We amplified mtDNA with a biotinylated version of the two primer pairs used by the FBI laboratory for DNA sequence analysis of the HVII region (25). Following amplification of the two fragments of the HVII region, the PCR products were mixed prior to the denaturation step and then hybridized to probes on the strips as described in Materials and Methods. The expected probe patterns were observed for each sample (data not shown).

The mtDNA typing strip in Fig. 1*A* is representative of results obtained from the extracted population database samples. Seventeen SSO probes are immobilized on each strip in a line format. The first probe on the strip is designated "S" and serves as an intensity control, analogous to the "C" probe on AmpliType HLA DQA1 probe strips. The "S" probe sequence is complementary to a region in HVII that appears to be relatively conserved across populations and between individuals according to our database queries described in Materials and Methods. Consequently, all amplified HVII products, regardless of their sequence in the remainder of the HVII region, can hybridize to this probe. The remaining 16 probes on the strip cover five polymorphic regions (defined by Stoneking et al. (12)) within the 415-bp amplified sequence. Polymorphisms are recognized by two SSO probes in region A, seven probes in region B, four probes in region C, two probes in region D, and a single probe in region E. The probe numbering scheme is explained in the legend to Table 2, which lists the sequence variants detected by the immobilized SSO probe panel.

As indicated in Table 2, the A-D region probes detect base substitutions between sequence variants. In contrast to regions A-D, the sequence variation detected in region E is a length polymorphism with a variable number of consecutive C residues. The number of consecutive C residues varies from 7 to 9 or more, with the majority of individuals carrying either 7 or 8. The region E probe was designed so that it does not hybridize to the "7C" sequence. Therefore, DNA from individuals whose mtDNA HVII sequence has seven C residues does not generate a probe signal in this region. Based on DNA sequence analysis, the presence of nine C residues or a mixture of sequences (e.g., 8 and 9 C residues) can lead to a weak E2 signal.

To interpret the mtDNA strip typing result, probe signals in regions A through E are compared to the "S" probe signal: probe signals darker than the "S" probe are considered positive. As indicated in Fig. 1*A*, an individual's "mitotype" is represented by five numbers corresponding to the positive probe signal observed in each of the five regions. For sample B216, a single positive signal was seen in each region: A2, B7, C4, D1, E2. When an individual's sequence is not complementary to any of the probes in a particular region, no probe signal is observed and the region is assigned a zero (e.g., region A of sample C147 and region D of sample H146 in Fig. 1*B*).

In addition to positive probe signals and "type 0" signals, two other types of signals may be observed within a region on the mtDNA probe strips. Occasionally, a sample yields a probe signal

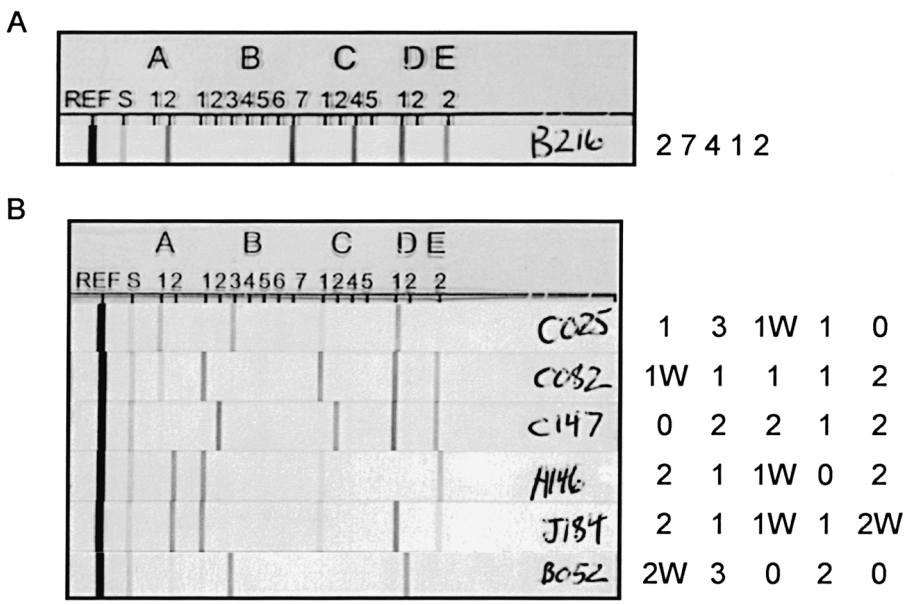


FIG. 1—*Representative immobilized SSO probe strip typing results:* (A) *The immobilized SSO probe strips are read using a template which indicates the locations of the 17 probes on the strip. The strip is correctly aligned on the template using the reference line at the left side of the strip ("ref"). Most often a single signal, greater in intensity than the S probe signal, will be observed in regions A-D; a positive signal in region E is observed in approximately 40% of the population. Individual B216 has a single, positive signal in each of the five HVII regions: A2, B7, C4, D1, E2. The mitotype is represented simply as 27412.* (B) *For any given sample, one or more regions on the strip may lack a signal. However, the proportion of samples that lack a signal within a region(s) is lower than the proportion of samples that yield a positive signal in each of regions A-D (see Table 4). Samples C025, C147, H146, and B052 have no signal in region E, A, D, and C/E, respectively. Regions that do not have a visible signal are designated with a "0" in the mitotype as indicated next to the strips. In rare instances (~5% of the tested samples), a signal that is visible but weaker than the intensity of the S probe signal will be observed. The weak signal may be similar in intensity to the S line, as with the weak A1 signal in sample C082, or it may be barely visible (see weak A2 signal in sample B052). Regardless of the relative intensity, a weak signal is designated with a "W" after the probe as indicated in the mitotypes for the five samples containing a weak signal.*

that is visible but lighter than positive signals in other regions on the same strip. Some of the strips in Fig. 1*B* are representative of weak probe signals observed in region A (C082 and B052) and region C (C025 and H146). All of the samples with weak signals in regions A-D were characterized by sequence analysis and, as expected, they each were found to contain a sequence variant that has a partially destabilizing mismatch to a specific probe (sequence analysis is discussed in detail in a later section). The fourth category of probe signal observed on the mtDNA probe strips is two signals within a single region (Fig. 2*A*, region B of sample B190). One potential source of additional probe signals is contamination, in which more than one individual contributes DNA to the sample. However, the presence of two signals in a single region might also be due to heteroplasmy, when an individual carries two detectable mtDNA sequences.

Population Variation of HVII

Using the immobilized SSO probe strips described above, we typed 689 individuals from four population groups (200 African American, 200 U.S. Caucasian, 200 U.S. Hispanic, and 89 Japanese individuals). A total of 142 different probe patterns was observed in the combined set of 689 samples when all four categories of probe signals were differentiated. The mitotypes observed in the four individual populations are listed in Table 3 along with their frequencies. Of the 82 mitotypes observed in African Americans, 41 of them were not observed in the other three populations. Similarly, 23 of the 62 mitotypes observed in U.S. Caucasians, 16 of 61 mitotypes observed in U.S. Hispanics, and 3 of 32 mitotypes observed in Japanese were not found outside of the respective populations. However, if the weak E2 signal is not distinguished from a positive E2 probe signal, then all of the mitotypes observed in the Japanese population were found in one or more of the other three population groups. The most frequent mitotype observed in the African American group was 24212 (8%) and was unique to this population group. In U.S. Caucasians, the two most frequent mitotypes were observed at approximately the same frequency (11112 at 11.5%; 21110 at 10.5%). The 11112 mitotype was seen in one African American and one U.S. Hispanic sample and was absent in the Japanese samples tested. However, the 21110 mitotype was the most frequent mitotype observed in the Japanese samples (15.7%) and was observed in 3 and 6% of the African American and U.S. Hispanic individuals, respectively. Additional comparisons of sequence variation among these and four other population groups (Somali, Bosnians, Finns, and Saami) have been made. Comparisons to published sequence data for HVII have also been carried out and these results are reported elsewhere (26).

The frequencies of the observed mitotypes were used to calculate the genetic diversity (h) values for each population using the following equation (27):

$$h = [1 - \sum (\text{mitotype freq})^2]\, n/(n - 1)$$

where n is the number of individuals in the database.

The genetic diversity values range from ~0.95 to 0.98 (Table 3). To compare the level of discrimination of this mtDNA typing system to nuclear markers commonly used for human identification applications, the genetic diversity (h) value can be compared to the power of discrimination (Pd) value ($Pd = 1 - \sum (\text{genotype freq})^2$). Genetic diversity values of ~0.95 to 0.98 indicate that the power of the immobilized SSO probe mtDNA typing system for distinguishing unrelated individuals is comparable to some of the original individual typing systems used for rapid sample screening. The HLA DQA1 and D1S80 markers both have Pd values in approximately the same range across these population groups. If the weak signals observed in region E are combined with corresponding mitotypes containing a normal E2 probe signal, the genetic diversity values decrease to ~0.92–0.98. If the region E probe is not used and the mitotypes consist of the probe signals observed in regions A-D, then the genetic diversity values decrease significantly to ~0.87–0.96.

Table 4 contains a list of the frequencies of the HVII sequence variants distinguished by the immobilized SSO probes in each region. The group designated "other" contains those samples that yielded weak signals or two signals within a single region. All five HVII regions are polymorphic in these four population groups with the exception of region A in the Japanese group, which is A2 exclusively. Type 0 sequences were observed in all population groups and in all HVII regions with the exception of region A. In region A, only one U.S. Caucasian individual had type 0; none of the African American, U.S. Hispanic, or Japanese individuals typed as A0. Sequence analysis of the HVII region was carried out on all samples yielding a weak signal or two signals within a single region and on 89 of the samples that contained type 0 in one or more regions. The sequence data are summarized in a later section.

Generally, the frequencies of the sequence variants detected by the immobilized SSO probe panel were similar to the frequencies of the corresponding HVII sequence variants detected by the SSO probes described by Stoneking et al. (12). However, there are some interesting exceptions in regions B and C between the African American population reported here and the African population used by Stoneking et al. (12). The observed frequencies of the C1 and C2 sequence variants were 17 and 33%, respectively, using the immobilized SSO probe panel and our African American population. Stoneking et al. (12) reported C1 and C2 frequencies of 36 and 19%, respectively. In region B, the B1 and B3 sequence variants were both observed at a frequency of ~17% in our African American population, while the B1 and B3 frequencies reported by Stoneking et al. (12) were 28 and 12%, respectively. Our 200 African American samples were collected from individuals from all regions of the United States. In contrast, the African population typed by Stoneking et al. (12) contained Nigerian samples (7 out of 129) and African American samples that may have been collected from regional samples that represent a particular subgroup (e.g., Afro-Caribbean).

The differences between the type 0 and reported "blank" frequencies are due to our adding probes in regions B and C; the variants distinguished by these new probes would type as "blanks" in Stoneking et al.'s dot blot system. The additional probes were designed to reduce the frequency of samples containing type 0 in regions B and C. Since African Americans had the highest frequency of type 0 sequences in both regions, we designed probes to distinguish sequence variants that were observed more frequently in that population group. As expected, the frequencies of the B4–B7 sequences and the C4 and C5 sequences were highest in the African American population (Table 4). Interestingly, the frequency of the B5 variant was nearly as high in the Japanese population as it was in the African American population.

Characterization of Type 0 and Weak Probe Signals Observed in the Population Database

As described above and listed in Table 2, SSO probes were designed to detect sequence variants in the five regions defined by Stoneking et al. (12). In addition to the sequence variants detected

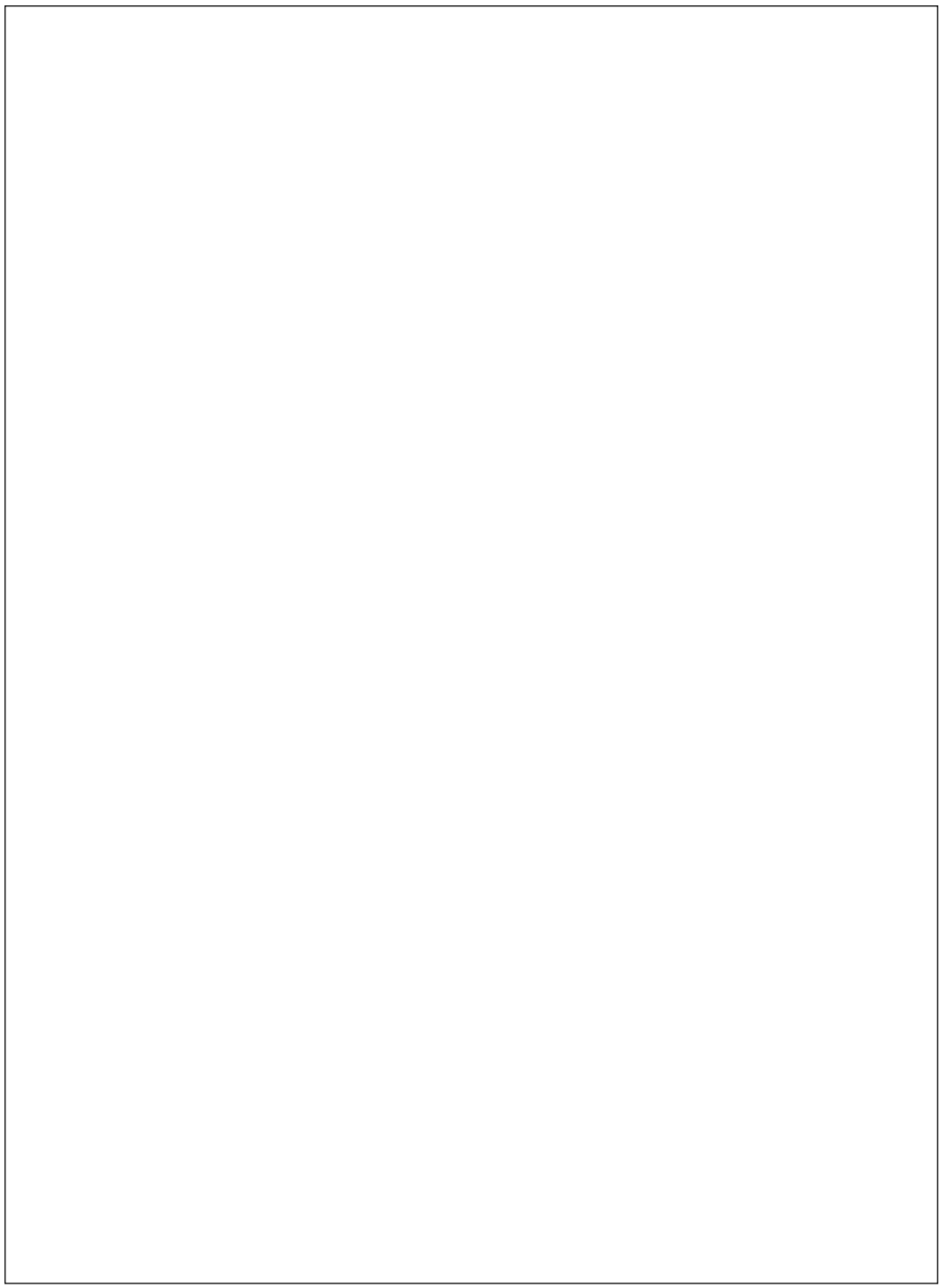


FIG. 2—*Detection of heteroplasmy by immobilized SSO probe typing and DNA sequence analysis. Individual B190 is heteroplasmic in region B as indicated by the two probe signals observed at the B3 and B4 probes on the strip shown in "A." The sequences distinguished by probes B3 and B4 differ at position 146: the B3 sequence has a T and the B4 sequence has a C at this position. B190 was sequenced using [^{33}P] dATP ("B") and fluorescent dye terminators ("C"). The sample on the right in "B" is B190 and the arrow indicates position 146 in the sequence. Both T and C are visible on the gel at that position. The sample on the left was included to show, for comparison, sequence containing only a T at position 146. The sequence obtained from sample B190 using the ABI Prism 310 Genetic Analyzer and Big Dye dye terminators is shown in "C." The arrow indicates position 146 (the numbers on the sequence printout are arbitrary units assigned by the software and do not correspond to the mtDNA sequence). Both T and C are visible in the sequence and the two signals were detected by the Factura software (note the "Y" in the sequence at the top of the peaks, which indicates the detection of both pyrimidine-containing bases at a single position in the sequence). Clearly, heteroplasmy can be detected by all analytical methods. However, some instances of heteroplasmy will not be detected by the immobilized SSO probe strips (see discussion in text) and will be detected only by DNA sequence analysis. The reverse situation, in which the immobilized SSO probe strips detect two sequences while DNA sequence analysis does not, has also been observed (see Fig. 3 and discussion for an example).*

TABLE 3—*Frequencies of mitotypes observed in each population group.*

AFRICAN AMERICAN h= 0.9795

	A	B	C	D	E	# OBS	FREQ
1	1	1	0	2	0	1	0.005
2	1	1	1	1	0	1	0.005
3	1	1	1	1	2	1	0.005
4	1	1	5	2	0	4	0.020
5	1	1	5	2	2	1	0.005
6	1	3	0	2	0	4	0.020
7	1	3	1	1	0	1	0.005
8	1	5	0	1	2	1	0.005
9	1	6	0	2	0	1	0.005
10	2	0	0	0	0	1	0.005
11	2	0	0	1	0	1	0.005
12	2	0	0	2	0	6	0.030
13	2	0	0	2	2	1	0.005
14	2	0	1	1	0	1	0.005
15	2	0	2	1	0	10	0.050
16	2	0	2W	1	0	2	0.010
17	2	0	2	1	2	8	0.040
18	2	0	2	1	2W	1	0.005
19	2	0	4	1	0	2	0.010
20	2	0	4	2	0	1	0.005
21	2	1	0	0	2	1	0.005
22	2	1	0	1	0	2	0.010
23	2	1	0	2	2	1	0.005
24	2	1	1	0	0	1	0.005
25	2	1	1	1	0	6	0.030
26	2	1	1	1	2	6	0.030
27	2	1	1	1	2W	1	0.005
28	2	1	2	1	0	1	0.005
29	2	1	2W	1	0	2	0.010
30	2	1	2	1	2	1	0.005
31	2	1	5	1	0	1	0.005
32	2	1	5	1	2	2	0.010
33	2	1	5	1	2W	1	0.005
34	2	2	0	1	0	1	0.005
35	2	3	0	1	2	1	0.005
36	2	3	0	2	0	9	0.045
37	2W	3	0	2	0	2	0.010
38	2	3	0	2	2	4	0.020
39	2W	3	0	2	2W	1	0.005
40	2	3	1	0	2	1	0.005
41	2	3	1	1	0	2	0.010
42	2	3	1	1	2	1	0.005
43	2	3	1W	2	0	1	0.005
44	2	3	2	2	0	6	0.030
45	2	3	2	2	2	2	0.010
46	2	4	0	1	0	1	0.005
47	2	4	0	2	2	1	0.005
48	2	4	1	1	0	2	0.010
49	2	4	2	1	0	5	0.025
50	2	4	2	1W	0	1	0.005
51	2	4	2	1	2	16	0.080
52	2	4	2W	1	2	1	0.005
53	2	4	2	1	2W	2	0.010
54	2	4	4	1	2	1	0.005
55	2	5	0	1	0	7	0.035
56	2	5	0	1	2	4	0.020
57	2	5	1	1	0	2	0.010
58	2	5	1	1	2	2	0.010
59	2	5	2	1	0	7	0.035
60	2	5	2	1	2	3	0.015
61	2	5	4	1	0	3	0.015
62	2	5	5	1	0	1	0.005
63	2	5	5W	1	0	1	0.005
64	2	5	5	1	2	2	0.010
65	2	5	5	1	2W	3	0.015
66	2	6	0	1	0	1	0.005
67	2	6	0	1	2	1	0.005
68	2	6	1	1	0	2	0.010
69	2	6	1	1	2W	1	0.005
70	2	6	2	1	0	2	0.010
71	2	6	2	1	2	1	0.005
72	2	6	4	1	0	4	0.020
73	2	6	4	1	2	2	0.010
74	2	6	5	1	0	1	0.005
75	2	6	5	1	2	1	0.005
76	2	7	0	0	0	1	0.005
77	2	7	1	1	2	1	0.005
78	2	7	4	1	0	4	0.020
79	2	7	4	1	2	4	0.020
80	2	7	4	1W	2	1	0.005
81	2	1,5	1	1	0	1	0.005
82	2	3,4	0	2	0	1	0.005
						200	1.000

U.S. CAUCASIAN h= 0.9564

	A	B	C	D	E	# OBS	FREQ
1	0	2	2	1	2	1	0.005
2	1	0	1	1	2	1	0.005
3	1	0	1	1	2W	2	0.010
4	1	1	0	1	0	1	0.005
5	1	1	1	1	0	17	0.085
6	1W	1	1	1	0	1	0.005
7	1	1	1	1	2	23	0.115
8	1W	1	1	1	2	4	0.020
9	1W	1	1	1	2W	2	0.010
10	1	1	1	1	2W	5	0.025
11	1	1	2	1	0	1	0.005
12	1	1	2	1	2	2	0.010
13	1W	1	2	1	2	1	0.005
14	1	2	1	1	0	3	0.015
15	1	2	1	1	2	1	0.005
16	1	2	1	1	2W	1	0.005
17	1	3	0	1	0	1	0.005
18	1	3	1	1	0	6	0.030
19	1	3	1W	1	0	1	0.005
20	1	3	1	1	2	5	0.025
21	2	0	0	1	0	1	0.005
22	2	0	1	1	2	1	0.005
23	2	0	2	1	0	2	0.010
24	2	0	2	1	2	2	0.010
25	2	1	0	0	0	1	0.005
26	2	1	0	1	0	13	0.065
27	2	1	0	1	2	3	0.015
28	2	1	1	0	0	1	0.005
29	2	1	1	1	0	21	0.105
30	2	1	1	1W	0	2	0.010
31	2	1	1	1	2	13	0.065
32	2	1W	1	1	2W	1	0.005
33	2	1	1	1	2W	2	0.010
34	2	1	1W	1	2W	1	0.005
35	2	1	2	1	0	7	0.035
36	2	1	2	1	2	2	0.010
37	2	1	2	1	2W	2	0.01
38	2	2	0	1	0	1	0.005
39	2	2	0	1	2	1	0.005
40	2	2	0	1	2W	1	0.005
41	2	2	1	1	0	1	0.005
42	2	2	2	1	2	2	0.010
43	2	3	0	0	0	1	0.005
44	2	3	0	0	2	1	0.005
45	2	3	0	1	0	1	0.005
46	2	3	0	1	2	1	0.005
47	2	3	1	1	0	3	0.015
48	2	3	1	1	2	4	0.020
49	2	3	2	1	2	2	0.010
50	2	4	0	1	0	1	0.005
51	2	4	1	1	0	3	0.015
52	2	4	1	1	2	5	0.025
53	2	4	2	1	0	1	0.005
54	2	5	0	1	2	1	0.005
55	2	5	1	1	0	4	0.020
56	2	5	2	1	0	2	0.010
57	2	5	2	1	2	1	0.005
58	2	6	1	1	0	1	0.005
59	2	6	1	1	2	3	0.015
60	2	6	2	1	0	2	0.010
61	2	7	1	1	2	1	0.005
62	2	1,3	1	1	2	2	0.010
						200	1.000

U.S. HISPANIC h= 0.9468

	A	B	C	D	E	# OBS	FREQ
1	1	1	0	2	0	1	0.005
2	1	1	1	1	0	6	0.030
3	1	1	1	1	2	1	0.005
4	1	1	1	1	2W	2	0.010
5	1	1	2	1	0	2	0.010
6	1	3	1	1	0	1	0.005
7	1	5	5	1	0	2	0.010
8	2	0	0	2	2	1	0.005
9	2	0	1	0	2	4	0.020
10	2	0	1	1	0	11	0.055
11	2	0	1	1	2	32	0.160
12	2	0	1	1	2W	11	0.055
13	2	0	2	0	2	1	0.005
14	2	0	2	1	0	1	0.005
15	2	1	0	0	0	2	0.010
16	2	1	0	0	2W	1	0.005
17	2	1	0	1	0	1	0.005
18	2	1	0	1	2	1	0.005
19	2	1	1	0	0	23	0.115
20	2	1	1	0	2	5	0.025
21	2	1	1W	0	2	1	0.005
22	2	1	1	0	2W	5	0.025
23	2	1	1	1	0	12	0.060
24	2	1	1	1W	0	1	0.005
25	2	1	1	1	2	9	0.045
26	2	1	1	1	2W	8	0.040
27	2	1	2	0	0	1	0.005
28	2	1	2	0	2	2	0.010
29	2	1	2	1	2	1	0.005
30	2	1	2	1	2W	2	0.010
31	2	1	5	1	2	1	0.005
32	2	2	0	1	2	1	0.005
33	2	2	0	1W	0	1	0.005
34	2	2	1	0	0	1	0.005
35	2	2	1	0	2	1	0.005
36	2	2	1	1	0	3	0.015
37	2	2	1	1	2	3	0.015
38	2	2	1	1	2W	1	0.005
39	2	3	0	2	2	1	0.005
40	2	3	1	0	0	2	0.010
41	2	3	1	0	2	1	0.005
42	2	3	1	1	0	2	0.010
43	2	3	1	1	2	2	0.010
44	2	3	1	1	2W	3	0.015
45	2	3	2	2	0	1	0.005
46	2	3	2	2	2W	1	0.005
47	2	4	1	1	0	2	0.010
48	2	4	1	1	2	3	0.015
49	2	4	1	1	2W	2	0.010
50	2	5	1	1	0	3	0.015
51	2	5	1	1	2	2	0.010
52	2	5	2	1	2	1	0.005
53	2	5	5	1	0	1	0.005
54	2	5	5	1	2	2	0.010
55	2	6	1	1	2	1	0.005
56	2	6	1	1	2W	1	0.005
57	2	6	4	1	0	1	0.005
58	2	6	4	1	2	2	0.010
59	2	7	0	1	2	1	0.005
60	2	7	2	1	0	1	0.005
61	2	3,4	0	0	2	1	0.005
						200	1.000

JAPANESE h= 0.9502

	A	B	C	D	E	# OBS	FREQ
1	2	0	1	1	0	1	0.011
2	2	0	1	1	2	1	0.011
3	2	1	0	1	0	3	0.034
4	2	1	0	1	2	4	0.045
5	2	1	0	1	2W	2	0.022
6	2	1	1	0	0	4	0.045
7	2	1	1	0	2	1	0.011
8	2	1	1	1	0	14	0.157
9	2	1	1	1	2	6	0.067
10	2	1	1W	1	2W	2	0.022
11	2	1	1	1	2W	6	0.067
12	2	1	2	1	2	1	0.011
13	2	1	2	1	2W	2	0.022
14	2	2	0	1	2W	1	0.011
15	2	2	1	0	0	1	0.011
16	2	2	1	0	2	1	0.011
17	2	2	1	1	0	1	0.011
18	2	2	1	1	2	2	0.022
19	2	2	1	1	2W	1	0.011
20	2	3	0	1	0	1	0.011
21	2	3	1	0	2	2	0.022
22	2	3	1	0	2W	1	0.011
23	2	3	1	1	0	6	0.067
24	2	3	1	1	2	5	0.056
25	2	4	0	1	0	1	0.011
26	2	4	1	1	2	2	0.022
27	2	5	0	1	0	3	0.034
28	2	5	0	1	2W	2	0.022
29	2	5	1	1	0	3	0.034
30	2	5	1	1	2	7	0.079
31	2	6	0	1	0	1	0.011
32	2	6	1	1	0	1	0.011
						89	1.000

TABLE 4—*HVII sequence variation frequencies listed as a percentage of the population.*

		AFAMER n=200	CAUC n=200	HISP n=200	JAPAN n=89	COMBINED n=689
IIA	1	7.5	35.0	7.5	0.0	14.5
	2	91.0	60.5	92.5	100.0	83.7
	0	0.0	0.5	0.0	0.0	0.1
	other*	1.5	4.0	0.0	0.0	1.6
IIB	1	17.0	62.5	44.0	50.6	42.4
	2	0.5	6.0	5.5	7.9	4.5
	3	17.5	13.0	7.0	16.9	13.1
	4	15.0	5.0	3.5	3.4	7.3
	5	18.0	4.0	5.5	16.9	10.2
	6	8.5	3.0	2.5	2.2	4.4
	7	5.5	0.5	1.0	0.0	2.0
	0	17.0	4.5	30.5	2.2	15.4
	other	1.0	1.5	0.5	0.0	0.9
IIC	1	16.5	69.5	82.0	74.2	58.3
	2	33.0	15.0	7.0	3.4	16.4
	4	11.0	0.0	1.5	0.0	3.6
	5	8.5	0.0	3.0	0.0	3.3
	0	27.5	14.5	6.0	20.2	16.5
	other	3.5	1.0	0.5	2.2	1.7
IID	1	73.0	97.0	71.0	88.8	81.4
	2	23.5	0.0	2.5	0.0	7.5
	0	2.5	2.0	25.5	11.2	10.2
	other	1.0	1.0	1.0	0.0	0.9
IIE	2	36.5	41.5	40.5	36.0	39.0
	0	58.5	50.0	41.0	44.9	49.2
	other	5.0	8.5	18.5	19.1	11.8

* The "other" category includes samples yielding weak or heteroplasmic signals. The frequencies of type 0 observed in the five HVII regions are comparable to the frequencies reported by Stoneking et al. (12) for corresponding population groups (type 0 is designated "blank" in their paper). The type 0 and "blank" frequencies can be compared directly in regions A and D because the same two sequence variants are distinguished by the two typing systems. To compare the type 0 and "blank" frequencies in region B, the frequencies of the B4-B7 sequence variants in Table 4 must be added to the type 0 frequency because the B4-B7 sequence variants type as "blank" using the probe panel described by Stoneking et al. (12). For region C, the frequencies of C4, C5 and type 0 in Table 4 must be added and the C3 and "blank" frequencies reported by Stoneking et al. (12) must be added together before the type 0 and "blank" frequencies can be compared. For region E, the type 0 frequency in Table 4 must be compared to the combined frequencies of E1 and "blank" types reported by Stoneking et al. (12).

TABLE 5—*Comparison of population variation parameters determined with two HVII probe sets.*

	AFRICAN AMERICAN		US CAUCASIAN		US HISPANIC		JAPANESE		COMBINED POPS	
n=	200		200		200		89		689	
HVII probe set	**EXPANDED**	**ORIGINAL**	**EXPANDED**	**ORIGINAL**	**EXPANDED**	**ORIGINAL**	**EXPANDED**	**ORIGINAL**	**EXPANDED**	**ORIGINAL**
h=	0.9795	0.9352	0.9564	0.9473	0.9468	0.9058	0.9502	0.9377	0.9788	0.9612
1 in	49	15	23	19	19	11	20	16	47	26
# unique mitotypes	82	50	62	50	61	51	32	29	142	112
% type 0 in:										
region B	17.0	64.5	4.5	17.0	30.5	43.0	2.2	11.0	15.4	39.3
region C	27.5	45.5	14.5	13.5	6.0	9.5	20.2	5.5	16.5	21.5

The "EXPANDED" probe set is the panel of 17 SSO probes immobilized on strips for the detection of sequence variation in HVII. The "ORIGINAL" probe set includes 11 of these 17 SSO probes and an additional probe corresponding to the HVII C3 sequence variant described by Stoneking et al. (12). All 689 samples were typed with the "EXPANDED" probe set and a probe for the C3 sequence. However, the C3 probe was not included in the final panel because it is not as discriminating as the combination of the new C4 and C5 probes. Samples containing the C3 sequence variant type as C0 in the new immobilized SSO probe set.

by their probes, four additional sequence variants in region B and two in region C are detected by the expanded HVII immobilized SSO panel described above. The sequence variant detected by their C3 probe is not detected by our panel. The addition of six new SSO probes had a modest effect on the genetic diversity value for HVII but dramatically reduced the number of samples that type as "0" in regions B and C (Table 5). The most significant decreases were observed in region B of African Americans (64.5 to 17%), U.S. Caucasians (17 to 4.5%), and Japanese (11 to 2%). The occurrence of type 0 in region C of African Americans was also greatly reduced (45.5 to 27.5%). The increase in the frequency of type 0 observed in region C of Japanese individuals is due to the elimination of the C3 probe. Eighty-nine samples containing a "0" in one or more regions (A-D) were sequenced; the results are summarized in a later section.

Probe signals that are visible but lighter than positive signals in other regions on the same strip are defined as "weak." Weak signals may be either comparable to or less than the S probe signal and examples of both levels of weak signals were observed (Table 6; Fig. 1*B*). Thirty of the 689 samples (~4%), spanning all four population groups, had a weak signal in one region (excluding region E). Weak signals were not observed at all probes; for example, of the seven region B probes, a weak signal was observed only at the B1 probe. Within populations, the frequency of weak signals was lower in U.S. Hispanics and Japanese (1.5 and 2.3%, respectively) than in U.S. Caucasians and African Americans (6.5 and 6.0%, respectively). These samples were sequenced to determine the sequence variation resulting in the weak signal; the results are summarized in a later section. Weak E2 signals were observed in 81 samples (~12%) and 27 of them were sequenced.

TABLE 6—*Summary of weak signal sequences.*

REGION	WEAK SIGNAL	SIGNAL INTENSITY	SEQUENCE VARIATION	# INDIVIDUALS BY POPULATION CAUC	AFAMER	HISP	JAPAN	TOTAL
A	1W	= S	72 C	8	-	-	-	8
	2W	< S	73 G 89 C 93 G 95 C	-	3	-	-	3
B	1W	< S	140 T	1	-	-	-	1
C	1W	< S	185 T	-	1	-	-	1
	1W	= S	204 C	2	-	1	2	5
	2W	< S	189 C 195 C	-	3	-	-	3
	2W	< S	195 C 204 C	-	2	-	-	2
	5W	< S	185 A 189 G 200 G	-	1	-	-	1
D	1W	< S	242 T	2	-	1	-	3
	1W	= S	244 G	-	1	-	-	1
	1W	< S	248 G	-	-	1	-	1
	1W	= S	257 G	-	1	-	-	1
TOTAL				13	12	3	2	30

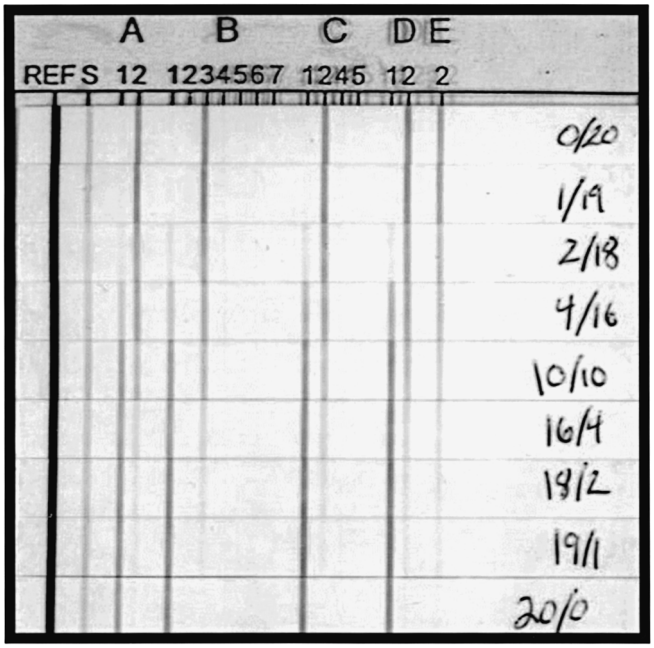
Detection of Mixtures of Sequences

Due to the matrilineal inheritance of the mitochondrial genome, it is expected that an individual will carry a single mtDNA sequence identical to his or her mother. The occurrence of two sequences in a sample may be due to contamination of the sample by a second source of mtDNA or to heteroplasmy. Heteroplasmy is the presence of more than one mtDNA sequence within an individual. There is increasing evidence that heteroplasmy within the control region is a more common occurrence than previously thought (28,29). When two sequences are present in a sample and are complementary to different probes in one of the HVII regions, two probe signals within the same region will be detected.

Of the 689 samples of DNA extracted from blood and typed with the immobilized SSO probe panel described above, five (<1%) appeared to have two sequences present in region B; multiple signals were not observed in the other regions in any of the 689 samples. These five samples were sequenced and, as expected, two bases were observed at one of the three positions distinguished by the region B probes (i.e., T and C at 146, 150, or 152). Multiple bases were not observed at any other position within the HVI and HVII regions, indicating that the secondary signals are unlikely to have originated from a contaminating DNA (the expectation being that two unrelated individuals are likely to differ at more than one position). A comparison of the detection of the two sequences by the immobilized SSO probe strips with both manual and automated sequencing is shown in Fig. 2 for one of the five heteroplasmic samples. Clearly, all methods were able to detect the major and minor sequences in this sample.

Because the ability to detect the presence of a second contributor or sequence is an important aspect of forensic DNA typing assays, we performed an additional study in which two DNA samples were mixed in varying ratios prior to amplification. The mitotypes of the two DNA samples are 23222 and 11110. The resulting PCR products were analyzed on immobilized SSO probe strips and the results are shown in Fig. 3*A*. The top and bottom strips correspond to the two DNA samples with the mitotypes described above. The results on the remaining strips were obtained from mixtures of these two samples in which the minor component comprised approximately 5, 10, or 20% of the total sample. The samples were also mixed 50:50. Regardless of which sequence is minor, the second signals in regions A-D are clearly detected when the minor sequence is present at ~10% of the total sample (see strips labeled 2/18 and 18/2). In region E, which has only one probe, the minor sequence is detected at the 10% level when it is E2 and the major sequence is E0. This level of detection is comparable to what many labs have reported for other immobilized SSO probe assays and D1S80 gels.

These samples were also analyzed with a fluorescent DNA sequencing system using Big Dye dye terminators and the 310 Genetic Analyzer following the manufacturer's protocols. PCR product from the same amplification reaction that was used to perform the strip typing was used for DNA sequence analysis. The initial volume of filtered PCR product added to the sequencing reaction contained an amount of product equivalent to what has



A

FIG. 3—*Detection of HVII sequence variation in mixtures of DNA samples of known sequence. Two DNA samples that differ in sequence in all five of the probe binding regions were mixed in defined ratios. Sequence variation in the mixed samples was detected by immobilized SSO probe analysis* (A) *and by DNA sequence analysis using the ABI Prism 310 Genetic Analyzer and Big Dye dye terminators* (B). *The strips and sequences labeled 0/20 and 20/0 represent the unmixed DNA samples. The two DNA samples were mixed so that each represented approximately 5% of the total DNA (1/19 and 19/1), 10% (2/18 and 18/2), 20% (4/16 and 16/4), and 50% (10/10). In the mixed samples, as the relative amount of one of the DNA samples is decreased (or increased), the corresponding probe signals on the immobilized SSO probe strips (A) also decrease (or increase) uniformly. The two DNA samples used to create the mixed samples have sequence differences in each of the five probe binding regions. These two samples also differ in sequence at positions 182 and 185. The portions of the sequence data that include these points of variation are shown in "B." The numbers at the top of the first set of boxes indicate the bases at which the two samples vary, and the arrows point to these specific positions in the sequences. The numbers on the sequence printout are arbitrary units assigned by the software and do not correspond to the reference mtDNA sequence (21). In contrast to the immobilized SSO probe strip results, the signals that define the sequence variation do not necessarily decrease or increase uniformly as the relative amount of one of the DNA samples in the mixed samples is varied (see text for discussion).*

given successful sequencing results in over 100 sequencing analyses of the HVII region. Three times (3×) this initial volume of filtered PCR product was also used and these sequencing reactions were set up in parallel with the 1× samples. While the peak heights for each position were higher with the 3× samples than the 1× samples, both sets gave the same results. Reactions containing five times the amount of PCR product were set up subsequently, but this amount was clearly an excess and created spurious small peaks at numerous positions, thus preventing interpretation. The sequences were examined both visually and using the Factura software included with the Genetic Analyzer and other Applied Biosystems automated sequencing instruments. The software was set to detect minor sequences that have a signal ≥15% of the major sequence signal.

Portions of the sequencing results obtained from the 3× samples are shown in Fig. 3*B*. Sequences at positions 73 (region A), 152 (region B), 195 (region C), and 247 (region D) are shown for the two reference samples (labeled 0/20 and 20/0 on the strips) and for the 2/18, 16/4, and 18/2 mixed samples. These two DNA samples also differ at positions 182 and 185 and the sequences at these positions are included in Fig. 3*B*. When the minor sequence is present at ~10% of the total sample (2/18 and 18/2), Factura detected two signals at positions 185 and 247. Very small peaks under the main peaks at positions 152 and 195 could be seen upon visual inspection but they are too small to be called with certainty. No trace of a second sequence was observed at position 73 by either inspection method. Similarly with the 18/2 sample, Factura did not detect a second sequence at any of these positions, but a very small peak under the main peak at position 185 was observed by visual inspection. If the sequence of the minor component had not been known, this peak may have been missed. These results are in contrast to the immobilized SSO probe strips which detected the minor sequence in all five regions in these two mixed samples. Even with the 16/4 sample in which the minor component is present at ~20% of the total sample, the Factura software set at ≥15% detected two sequences only at positions 73 and 182. The second sequence at the other positions was detected by visual inspection, but the peaks at positions 152 and 247 were too small to be called with certainty. As mentioned above, detection of the minor component was not enhanced by using a higher amount of PCR product in the sequencing reaction.

Sequencing with AmpliTaq FS DNA polymerase and Big Dye dye terminators, or any other combination of sequencing enzymes and nucleotides, produces a characteristic signal at each position in the sequence. These signals are reproducible but not equivalent and can be affected by changes at neighboring bases (e.g., note that the relative height of 74T when the sequence contains a G at 73 is increased significantly when the sequence contains an A at 73). At any given position, the substitution of one base for another can lead to a signal of different height, again characteristic for that base at that position. By comparing peak heights in the 0/20 and 20/0 samples at the positions that differ between these two sequences, it is readily apparent that the heights of peaks at a given position can differ by two-fold or more (e.g., 73G in 0/20 and 73A in 20/0). Use of different primers and/or different fluorescent sequencing chemistries (e.g., rhodamine dye) may produce more uniform peaks. However, the relative amounts of two sequences in a sample might not be determined reliably with any sequencing system unless a panel of mixtures containing the two sequences in a variety of known ratios is sequenced. Unfortunately, for most situations a supply of DNA samples with the sequences in question will not be available for performing the mixed sequence study. Therefore, attempts to deduce definitively the mtDNA sequence profiles of the contributors to samples containing a mixture of DNAs should be undertaken with caution. When interpreting data that apparently involve two sequences, it is important to consider that the results may not reflect the complete profile of the minor contributor when the relative amounts are significantly different (e.g., 1:10). When the amounts of DNA from two contributors are more similar and the sequences differ at multiple positions, many combinations of two sequences are possible.

Effect of Sequence Variation on HVII SSO Probe Signals

Individuals who have the same mitotype which includes a "0" in a particular region may have different sequences that prevent binding to the probes in that region. Similarly, weak signals may result from more than one sequence variant. The 30 samples that yielded weak signals and 89 samples that typed as "0" in one or more of regions A-D were sequenced. Tables 6 (weak signals) and 7 (type

B

FIG. 3—*(continued)*

TABLE 7—*Summary of type 0 sequence variation.*

REGION	SEQUENCE GROUP	SEQUENCE VARIATION	MOST SIMILAR TO*	# INDIVIDUALS BY POPULATION				TOTAL
				CAUC	AFAMER	HISP	JAPAN	
A	1	72 G	A1	1	-	-	-	1
B	1	143 A 146 C 152 C	B4	1	5	1	-	7
	2	143 A 152 C	B3	1	1	-	-	2
	3	146 C 150 T	B2/B5	-	1	-	-	1
	4	146 C 152 C 153 G	B4	-	-	1	-	1
	5	146 C 153 G	B2	-	-	9	-	9
	6	151 T 152 C	B3	-	2	1	-	3
	7	153 G	B1	5	1	1	1	8
C	1	185 A	C1	5	1	1	-	7
	2	185 A 188 G	C1	2	-	-	-	2
	3	185 C 195 C 198 T	C4	-	-	1	-	1
	4	186 A 189 C	C1	-	1	-	-	1
	5	186 A 189 C 195 C	C2	-	3	1	-	4
	6	186 T 195 C 200G	C2	-	-	-	1	1
	7	189 G	C1	-	3	-	-	3
	8	189 G 195 C 204 C 207 A	C2	3	-	-	-	3
	9	189 G 203 A	C1	-	-	-	1	1
	10	189 G 204 C 207 A	C1	-	1	-	-	1
	11	194 T	C1	-	-	1	4	5
	12	194 T 195 C 198 T	C4	-	1	-	-	1
	13	198 T	C1	-	-	1	-	1
	14	199 C	C1	-	-	1	-	1
	15	199C 203A	C1	1	-	-	-	1
	16	199 C 204 C	C1	-	2	-	-	2
	17	199 C 204 C 207 A	C1	1	-	-	1	2
	18	199 C 207 A	C1	-	-	1	-	1
	19	200 G	C1/C5	-	3	-	-	3
D	1	Δ249	D1	-	2	5	5	12
	2	246C	D1	-	-	1	-	1
	3	250 C	D1	2	1	-	-	3
			TOTAL	22	28	26	13	89

* based on the nucleotides at position 73 (region A), positions 146, 150, 152 (region B), positions 189, 195, 198, 200 (region C), or position 247 (region D)

0) contain summaries of the sequencing results. The sequence variation resulting in weak or type 0 signals are discussed here by region.

In region A, only one individual typed as "0" and the sequence analysis revealed a T to G transversion at position 72 (see C147 in Fig. 1*B*). Of the 11 individuals who had a weak signal in region A, eight of them (all Caucasian) had a T to C transition at position 72 and a signal at the A1 probe comparable in intensity to the S probe signal (e.g., see C082 in Fig. 1*B*). This difference in signal intensities (type "0" versus weak) between these position 72 variants is due to the different mismatches created between the PCR product and the A1 probe. The A1 probe is complementary to the antisense strand of the HVII sequence and contributes the T of the A:T pair that forms at position 72 between the A1 probe and the HVII PCR product (see Table 1). If position 72 is changed from a T to a G, then a C-T mismatch will be created (T from the A1 probe and C from the antisense strand of the PCR product). If instead position 72 is changed from a T to a C, then a G-T mismatch will be created (T from the A1 probe and G from the antisense strand of the PCR product). The C-T mismatch created by the 72G variant is more destabilizing than the G-T mismatch created by the 72C variant and results in a type 0 signal.

The remaining three individuals who had a weak signal in region A (A2 < S probe signal; e.g., see B052 in Fig. 1*B*) unexpectedly showed no sequence variation within the probe binding region. However, these three individuals shared the following sequence variation which was not observed in any of the remaining 126 samples that were sequenced: 89C, 93G, 95C. A possible explanation for this observation is that when these three sequence variations are present with 73G (A2), a somewhat stable secondary structure can form that greatly reduces the availability of PCR product for probe binding. The 89C variation may be sufficient for stabilizing this structure because an individual with 73G, 93G, and 95C typed as a normal A2 as did two individuals with 73G and 93G.

In region B, only one individual had a weak signal (B1 < S probe signal; see C168 in Fig. 4*A*). In this sample, a C to T transition at position 140 creates a G-T mismatch at one end of the hybridized probe and PCR product. Thirty-one of the 106 individuals who typed as "0" in the B region were sequenced. These individuals fell into one of seven sequence groups (Table 7) and were present in all four populations. The sequences defining the groups differ by only a single base from one or two of the region B sequences listed in Table 2. These single base differences create destabilizing mismatches that prevent binding of amplified product to the corresponding probe. The sequence variations in groups 1, 2, and 7 create a G-T mismatch either 3 or 8 bases from one end while in groups 3–6 there is a C-A mismatch at positions 6 to 12 bases from one end. These sequence variations are consistent with the observed type 0.

In region C, 12 individuals had a weak signal at the C1, C2, or C5 probe (Table 6). The weak C1 signals fell into two sequence groups. In one group, which contains a single individual, the sample has a barely visible C1 signal, while samples in the other group yield a C1 signal equal to the S probe signal (compare normal C1 signal in C168 to other samples in Fig. 4*A*). The sequence variation in the first group creates a C-T mismatch at one end of the hybridized probe and PCR product and yields a signal lighter than the S probe signal (see B075 in Fig. 4*A*). The second group has a C-A mismatch one base from one end and a signal equal to the S probe signal (see C025 and J184 in Fig. 4*A*). The individuals in the two sequence groups for the weak C2 signals have either a G-A or G-T

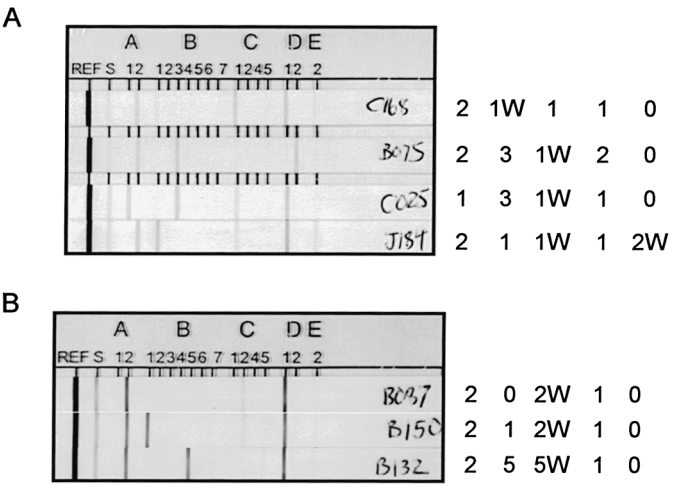


FIG. 4—*Immobilized SSO probe strip results from samples containing a weak signal in region C. Weak signals were observed at probes 1, 2, and 5 in the C region. Sample B075 in "A" yields a barely visible C1 signal that has a much lower intensity than the S probe signal. In contrast, samples C025 and J184 have a different sequence and yield a weak C1 signal that is comparable to the S probe signal (see Table 6 for the two sequence groups for the C1W signals). SSO strips from samples yielding a weak probe signal at either C2 or C5 are shown in "B." Weak signals at the C2 probe were divided into two sequence groups (Table 6). Sample B150 is in the first group and B087 is in the second group and both samples yield weak C2 signals that are less intense than the S probe signal. The only sample in the database to yield a weak C5 signal is B132, and the C5 signal is much lighter than the S probe signal.*

mismatch three bases from one end of the hybridized probe and PCR product. Their C2 signals are lighter than the S probe signal (see B087 and B150 in Fig. 4*B*). The individual who typed as a weak C5 (signal much lighter than the S probe signal) has sequence variation that creates a C-T mismatch at one end of the hybridized probe and product (see B132 in Fig. 4*B*). Forty-one individuals who typed as C0 were sequenced and divided into 19 sequence groups. In general, the mismatch created by the sequence variation in each group was highly destabilizing and/or located approximately in the middle of the probe.

In region D, six individuals had a weak D1 signal (Table 6); none of the weak D1 signals were observed in the Japanese population. These six individuals fell into one of four sequence groups, two of which generated a D1 signal equal to the S probe signal and two of which generated a D1 signal less than the S probe signal. All of the mismatches are G-T and are located between 3 and 12 bases from one end of the hybrid. Sequences of the 16 samples in the database with type 0 in region D fell into 3 groups (Table 7). The most common sequence variant, Δ249, was not observed in Caucasians. The sequence variation in the other two groups creates a C-A mismatch with the D1 probe that is located either 10 or 11 bases from one end of the probe, again consistent with the observed type 0. Figure 5 shows samples representing normal and weak D1 signals and type 0 in the D region. No weak D2 signals were observed.

Region E was sequenced in 62 samples to characterize the E2 probe signals that are designated type 0 (e.g., C025 and B052 in Fig. 1*B*), normal (e.g., C082, C147, and H146 in Fig. 1*B*), and weak (e.g., J184 in Fig. 1*B*). All 20 of the type 0 samples contained 7C residues and all 14 of the samples yielding a normal E2 probe signal contained 8C residues (Table 8). The remaining 28 samples selected for sequence analysis fell into nine sequence categories

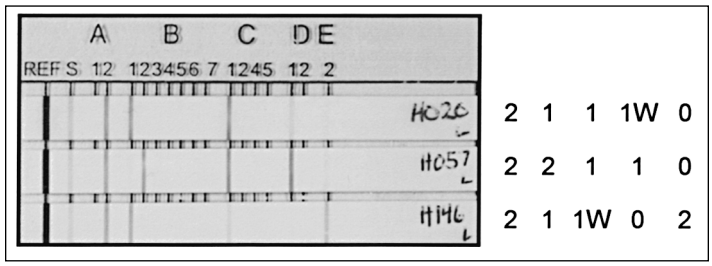


FIG. 5—*Immobilized SSO probe strip results from samples containing a weak signal or type 0 in region D. Weak signals were observed in six individuals at the D1 probe; no individual in the database has a weak D2 probe signal. Sample H057 has a normal positive D1 probe signal and H020 has a weak D1 probe signal that is lighter than the S probe signal. Sample H020 is in the first sequence group for weak D1 signals in Table 6. Sample H146 types as "0" in region D and carries the Δ249 sequence variation.*

TABLE 8—*Summary of sequence variation in region E.*

SIGNAL	GROUP	SEQUENCE VARIATION	# INDIVIDUALS BY POPULATION AFAMER	CAUC	HISP	JAPAN	TOTAL SEQ'D	TABLE 9 CODE 309.1	309.2	309.3
E0	1	7C	3	11	4	2	20	*	.	.
E2	1	8C	3	3	6	2	14	C	.	.
E2W	1	9C	-	-	-	1	1	C	C	.
	2	7C>8C	-	3	-	-	3	c	.	.
	3	7C<8	-	-	1	-	1	C+	.	.
	4	8C=9C	-	1	1	-	2	C+	C+	.
	5	8C<9C	3	6	4	1	14	c	C	.
	6	8C>9C	-	1	-	1	2	C	c	.
	7	9C=10C	-	-	1	1	2	C	C+	C+
	8	9C>10C	-	-	1	1	2	C	C	c
	9	8C<9C>10C	-	-	-	1	1	c	C	c
TOTAL			9	25	18	10	62			

that yield a weak E2 signal. One of these categories contains a single sample that has 9C residues, while the remaining samples in the other eight categories contain a mixture of sequences with different lengths of C residues (Table 8). The majority of individuals have either 7C residues (339 out of 689 or 49%) or 8C residues (269 out of 689 or 39%). The remaining 81 samples in our database (~12%) have either 9C residues or a mixture of different lengths of C residues.

Summary of DNA Sequence Analysis

DNA sequence analysis was performed on 20% of the African American, U.S. Caucasian, and Japanese samples and on 15% of the U.S. Hispanic samples. The HVII sequences obtained from these 129 individuals (42 African Americans, 40 U.S. Caucasians, 29 U.S. Hispanics, 18 Japanese) are summarized in Table 9. Sixteen of the observed variants have not been reported to Mitomap (30), and they are listed in Table 10.

As discussed in a previous section, two probe signals were observed in region B in five samples. Sequence analysis of both strands of these samples confirmed this observation. Of the remaining 124 samples that were sequenced, 123 do not appear to have a mixture of two sequences at any position outside of the variable C-stretch in HVII. One of the 129 samples (No. 98 in Table 9) was sequenced because it showed two probe signals in the C region: C1 and a weaker C3. Sequence analysis revealed both T (C1 signal) and a weaker C (C3 signal) at position 199. However, the C3 probe from the original set defined by Stoneking et al. (12) is not included in our expanded set (see Table 4). Therefore, heteroplasmy at this position will not be detected by this immobilized SSO probe panel.

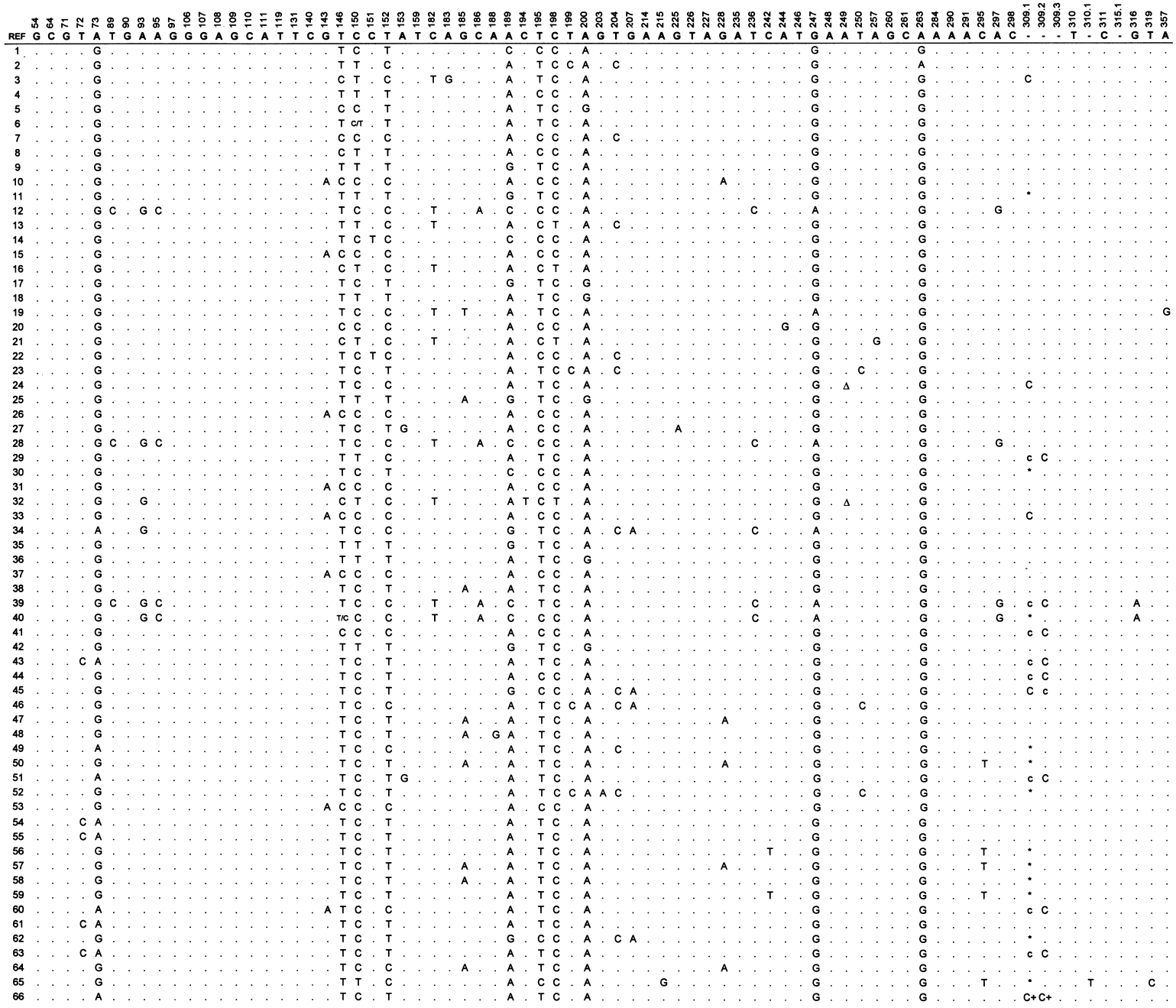
Casework Application

Shed hairs are among the most frequently collected types of evidence. These hairs typically lack a visible root and usually cannot be typed using PCR-based nuclear marker typing systems (31). Traditionally, shed hairs are examined microscopically and compared to 20 or more reference hairs from each suspect or victim. This method, which requires specialized training, is quite subjective and it is not possible to assign probabilities of identity when an evidentiary hair is considered to be a match with a particular individual. Therefore, development of a method for obtaining more definitive information from shed hairs would be very valuable. Mitochondrial DNA analysis is ideal for evaluating hairs that lack visible roots because mtDNA can be extracted from the root end and shaft portions of these hairs.

To illustrate the value of mtDNA analysis of shed hairs, using both the immobilized SSO probe strips and DNA sequencing, we describe a case involving multiple bombings at various sports facilities in Sweden. Two suspects were identified in connection with these bombings, which appeared to be carried out by the same person or group. Suspect 1 is a good friend of suspect 2, who admitted involvement in the bombings; suspect 1 denied all involvement. A bomb was found in a garage owned by suspect 1's mother and eight shed hairs were collected from tape and materials wrapped around the bomb. These hairs, along with plucked hairs from the two suspects, were typed using the immobilized probe strips (Fig. 6). The mitotypes for the eight hairs are listed in Table 11. One of the hairs (1-4) has the same mitotype as suspect 2 and two of the hairs (1-5 and 6-1) have the same mitotype as suspect 1. The remaining five hairs have mitotypes that differ in one or more regions from the two suspects, thus excluding the suspects as donors of these hairs. The 21112 mitotype observed in one of the hairs and suspect 2 is found in approximately 3–6.5% of the U.S. population groups (Table 3). The 25112 mitotype, found in two of the evidentiary hairs and in suspect 1, has been observed in 1% of African Americans and U.S. Hispanics, but not in any of our 200 U.S. Caucasian individuals (Table 3).

Reference hairs and the three shed hairs that matched the suspects (1-4, 1-5, and 6-1) were sequenced in both the HVI and HVII regions to try to further reduce the probability of identity by chance. The DNA sequencing results are shown in Table 12. Hair 1-4 differs from the reference hair of suspect 2 at three positions within HVII that are not detected by the immobilized SSO probes (228, 295, and 325). There are four positions within HVI at which the evidence hair differs from the suspect 2 reference hair (16069, 16126, 16256, and 16270). Based on these sequence differences, suspect 2 can be eliminated as the donor of this hair. Hairs 1-5 and 6-1 have the same sequence in both HVI and HVII as the reference hair from

TABLE 9—*HVII sequence variation in 129 individuals.*



Individuals 1-42 are African American, 43-82 are US Caucasians, 83-111 are US Hispanics, and 112-129 are Japanese. The positions distinguished by the immobilized SSO probe panel are included for each individual. The code for the C-stretch (309.1, 309.2, 309.3) is in the right-hand column of Table 8. For the heteroplasmic samples, if the two sequences were not present in approximately equal amounts the minor sequence is represented by a lower-case letter.

TABLE 10—*List of unique and rare sequence variants not reported in Mitomap (27). These sequences were observed in one African American (sample 8 in table), six U.S. Caucasians (samples 3, 5, 7, 9, 12, 14), and four Japanese (samples 6, 11, 13, 16). Variants were observed in five U.S. Hispanic individuals. Three of them carried a single variant (samples 10, 15, 16) but one U.S. Hispanic individual carried the 54T, 71 deletion, and 290-291 deletion variants and another carried the 90A and 106-111 deletion variants. One Japanese individual and two U.S. Hispanic individuals carried the 290-291 deletion variant.*

	Position	Variation	Reference
1	54	T	G
2	71	Δ	G
3	72	G	T
4	90	A	G
5	119	C	T
6	131	C	T
7	140	T	C
8	183	G	A
9	228	T	G
10	246	C	T
11	284	G	A
12	295	A	C
13	298	T	C
14	310	T INS	…
15	106–111	Δ	GGAGCA
16	290–291	Δ	AA

TABLE 11—*Mitotypes of evidence hairs and reference samples. Eight shed hairs were collected from materials wrapped around a bomb. These shed hairs and reference samples collected from the two suspects were extracted and typed using the immobilized SSO probe strips.*

Sample	A	B	C	D	E
Hair 1–4	2	1	1	1	2
Hair 1–5	2	5	1	1	2
Hair 5–1	2	0	2	1	0
Hair 5–2	1	1	1	1	0
Hair 5–3	2	5	1	1	0
Hair 5–4	1	3	1	1	2
Hair 6–1	2	5	1	1	2
Hair 1–2	2	5	0	1	0
SUSPECT 1 REF SAMPLE	2	5	1	1	2
SUSPECT 2 REF SAMPLE	2	1	1	1	2

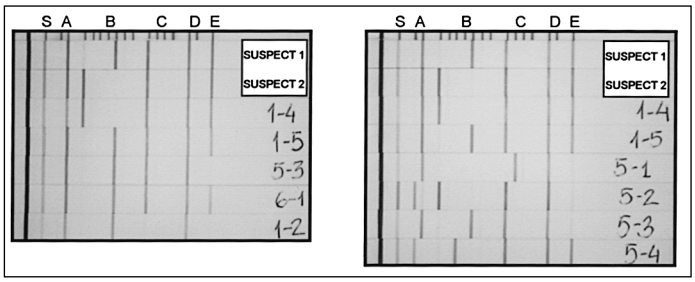


FIG. 6—*Immobilized SSO probe strip results obtained from shed hairs and reference samples from two suspects in a bombing case in Sweden. Eight shed hairs were collected from materials wrapped around a bomb. The hairs were extracted and then typed in duplicate on two separate occasions. The two sets of strips shown in this figure include at least one SSO probe strip result for each hair and a second result for three of the hairs (remaining strips not shown in this figure). The mitotypes for the hairs and reference samples are listed in Table 11. On some of the SSO probe strips, additional weak signals may be visible. These signals are most likely due to the use of excess PCR product during the strip typing procedure.*

suspect 1. Therefore, suspect 1 cannot be excluded as the donor of hairs 1-5 and 6-1. Hair 5-3 was also sequenced because it differed from hairs 1-5 and 6-1 only in region E. The frequency of heteroplasmy in the stretch of C residues in region E is higher than the frequency observed in nucleotides in the other regions. In addition, a single hair from a heteroplasmic individual may carry only the minor sequence which may or may not be detectable in other hairs or blood (Reynolds et al., in preparation; 32). Therefore, while unlikely, it is possible that suspect 1 is heteroplasmic in region E at an undetectable level and that this hair originated from suspect 1 but carries only the minor region E sequence. DNA sequence analysis of HVI revealed that hair 5-3 differs from hairs 1-5, 6-1, and suspect 1 at seven positions. Therefore, hair 5-3 clearly originated from another individual. In summary, none of the hairs removed from the bomb could be associated with suspect 2 who admitted involvement in the bombings. However, two of the hairs could have originated from suspect 1 who denied all involvement in these incidents.

This case illustrates the utility of the immobilized SSO strips as a screening tool. Out of eight evidentiary hairs, half were eliminated immediately and unambiguously by the strip typing results. Consequently, only four of the eight hairs needed to be sequenced. In two other cases involving shed hair evidence, eight out of 12 hairs in one case and four out of four hairs in the other case were excluded by the strips. Using the strips to screen the evidentiary hairs greatly reduced the time required for resolving these two cases (1 day versus several days). The two additional hair cases and others will be discussed in detail elsewhere (M. Allen et al., in preparation).

TABLE 12—*Sequence variation between shed hairs and reference samples. The HVI and HVII sequences of four shed hairs were compared to reference samples collected from the two suspects. The reference sequence in bold-face type follows the numbering scheme in Anderson et al. (18).*

		HVII						HVI							
		150	228	295	309	325		16069	16126	16144	16148	16256	16261	16270	16311
Sample	ref	C	G	C	C	C	ref	C	T	T	C	C	C	C	T
HAIR 1-4		C	G	C	C	C		C	T	T	C	T	C	T	C
HAIR 1-5		T	G	C	C	C		C	T	C	T	C	C	C	C
HAIR 5-3		T	G	C	—	C		C	T	T	C	T	T	C	T
HAIR 6-1		T	G	C	C	C		C	T	C	T	C	C	C	C
SUSPECT 1 REF SAMPLE		T	G	C	C	C		C	T	C	T	C	C	C	C
SUSPECT 2 REF SAMPLE		C	A	T	C	T		T	C	T	C	T	C	C	C

Conclusions

Based on the population data collected from 689 unrelated individuals and the DNA sequence analysis of 129 samples, we conclude that the HVII immobilized SSO probe typing system provides valuable, discriminating information from samples that do not yield results with nuclear DNA-based systems. It is particularly useful for screening large numbers of hairs and bloodstains. As with any forensic DNA typing test, careful consideration must be paid to how results are interpreted and reported. Mitochondrial DNA analysis involves some unique interpretation issues which are discussed below.

Interpretation of the mitotypes obtained from the immobilized SSO probe strips is straightforward, particularly when a single signal greater than the S probe signal or no signal is observed in each region. Two rare types of results within a region, two probe signals due to heteroplasmy and weak signals, require more careful interpretation.

When more than one region has more than one probe signal, there may be a second or third contributor to the sample. The presence of two signals in only one of the regions may be due to a second contributor, but may also be due to heteroplasmy. Based on the samples in our population database that had two signals within one region and our work with shed hairs and tissue samples, the region most likely to show two sequences as a result of heteroplasmy is region B. Heteroplasmy is observed at a higher frequency in region E, which includes the variable C-stretch, but it will not be detected as two signals because there is only a single probe for region E. Heteroplasmy in the C-stretch of region E may appear as a weak E2 signal.

Two probe signals within a single region for a given sample are consistent with either heteroplasmy or multiple contributors. The additional sequences observed in region B of five samples in the database are not likely due to the presence of multiple contributors (i.e., contamination) rather than heteroplasmy. First, the HVI region of these samples was sequenced. Each of these samples had between two and ten differences from the reference sequence (21). Multiple bases were not observed at any of these positions, supporting the conclusion that these samples are not contaminated. In addition, possible contamination of the amplification reaction during set-up can be ruled out because the same mtDNA result was obtained from separate amplifications of the samples performed by different people over a two-year period. Also, the mitotypes of these five samples are not consistent with the individuals who performed the amplification reactions.

While observing heteroplasmy in only five out of 689 blood samples (0.7%; 95% confidence interval 0.02–1.7%) suggests that the frequency of heteroplasmy in the HVII region is very low in the general population, this value is likely to be an underestimate of the overall frequency of heteroplasmy for the following reasons. First, most instances of heteroplasmy may not be revealed in blood samples. In a study in which over 400 hairs from 24 people were typed in the HVII region, eight (33%) of the individuals had one or more heteroplasmic hairs in regions A-D (Reynolds et al., in preparation). When the E region C-stretch is included, 14 (61%) of the individuals had one or more heteroplasmic hairs in regions A-E. Five to 20 hairs were typed from 17 individuals and 21–60 hairs were typed from the remaining seven individuals. Heteroplasmic hairs were observed within the first ten hairs tested from the eight individuals who are heteroplasmic in one of the non-E regions and within the first five hairs tested for four of those individuals. HVII typing of blood samples from these 24 individuals did not reveal heteroplasmy, even in the eight individuals who are heteroplasmic in regions A-D with the following exception. An extremely light probe signal was visible at a second probe in region B for one of these eight individuals but no additional base was detected by DNA sequence analysis. Four of the eight heteroplasmic individuals had at least one hair with a different dominant sequence than the blood sample and three hairs appeared to have a complete switch in sequence. In these instances, the difference was observed at a single position and further sequence analysis and testing confirmed that the sequence originated from the individual in the study, not from contamination (Reynolds et al., in preparation).

The second reason that the frequency of 0.7% (95% confidence interval 0.02–1.7%) might be an underestimate, regardless of the type of sample being analyzed, is that not all sequence variants are detected by this panel of probes (e.g., the sample with both T and C at position 199). Obviously, sequence variation outside the probe binding regions will not be detected. In addition, if the second sequence in a heteroplasmic sample has a highly destabilizing mismatch in a probe-binding region, it will not be detected by the strips and the sample will appear homoplasmic, possibly with a weaker probe signal for the major sequence. Similarly, even if each of the two sequences can bind to its corresponding probe, the minor sequence may not be detected if it is present at less than ~10% of the total PCR product. However, as discussed earlier, sequence analysis of the HVII region from 123 apparently homoplasmic individuals did not reveal additional heteroplasmic positions.

As shown in Figs. 1, 4, and 5, weak signals can be barely visible (e.g., region C of B075 in Fig. 4) or clearly visible and similar in intensity to the S line (e.g., region A of C082 in Fig. 1*B*). These signals are characteristic of an individual's HVII sequence and are typed reproducibly on the immobilized SSO probe strips, even from separate amplification reactions. When both evidentiary material and a reference sample have the same mitotype with the weak signal, the donor of the reference sample cannot be excluded as the donor of the evidentiary material and the samples should be sequenced.

Other possible causes of a weak signal in a single region are multiple contributors to the sample and heteroplasmy. However, only some combinations of two sequences due to contamination or heteroplasmy will produce a weak signal on the strip in a single region. For example, the major sequence in a sample must be type 0 in one region and the minor sequence must type positively in this region and have the same type as the major contributor in all other regions. With mixed samples containing mtDNA from two individuals, multiple signals in one or more regions or unequal signal intensities across regions are more likely to be observed than a weak probe signal in a single region.

Whether a sample yields a weak signal because of sequence variation or the presence of multiple sequences (i.e., contamination or heteroplasmy) can be determined unequivocally by DNA sequence analysis. The sequencing results will usually reveal either a single sequence containing variation that affects probe binding or, when two different sequences are present, two bases at a position within or near the region yielding the weak probe signal. However, to distinguish between mixed samples due to contamination and heteroplasmy as the source of the two sequences, the analyst must consider the number of nucleotide differences between the two sequences, the position of the variable base(s) within the control region, the type of biological material tested, the method of analysis, and the results obtained from controls and reference samples.

Based on our data collected from 689 blood samples, detectable

heteroplasmy is very rare (0.7%; 95% confidence interval 0.02–1.7%) and heteroplasmy has been detected by the probe set only in the region B sequences. These observations aid the interpretation of results obtained from various reference and evidentiary blood samples (hair samples will be discussed separately). For example, if a suspect's reference blood sample mitotype is 12110 and the evidentiary blood sample's mitotype is 22110, the result is considered inconclusive according to guidelines presented at various forensics meetings because the single base difference at position 73 could be due to heteroplasmy, even when there is no evidence of heteroplasmy in other samples. Our data indicate that it is much more likely that these two blood samples originated from different individuals than from a single heteroplasmic individual. Also, some probe pairs (e.g., B1 and B4) distinguish sequences that differ at two positions. Blood samples with mitotypes differing only at these probes would be even more likely to have come from different individuals. In addition, there is no known published evidence that blood will yield a different mitotype when sampled over an individual's lifetime, although pre- and post-transplant blood samples from a bone marrow transplant recipient may yield a different mitotype. As always, the specific details of each case should be considered before making final conclusions.

When analyzing individual hairs rather than blood samples, more testing must be done before similar conclusions can be made. In a heteroplasmic person, it is possible for an individual hair to carry a different sequence than the sequence obtained from the person's blood (Reynolds et al. in preparation; 32). Sometimes an individual will not appear to be heteroplasmic when a reference blood sample is typed because the minor secondary sequence is not present at a detectable level. However, one hair from this heteroplasmic individual may carry the same primary sequence as the blood while a different hair may carry this sequence and a second detectable sequence. Yet another hair from this individual may have only one detectable sequence but it will be the secondary sequence. Therefore, it is more relevant to compare evidentiary hairs to multiple individual (i.e., unpooled) hairs from a suspect than to a reference blood sample, particularly when the types obtained from the evidentiary hairs are consistent with heteroplasmy. However, a reference blood sample could be analyzed in addition to the reference hairs to confirm the individual's dominant sequence.

In conclusion, the immobilized SSO probe approach provides a very rapid method for detecting sequence variation in HVII: following amplification, typing of up to 40 samples takes 2–2.5 h. Another valuable feature of this typing method is that the analyst can see immediately if two or more samples have the same or different mitotypes. Even with automated sequencing approaches, two or more days are required to sequence and evaluate samples because multiple sequencing reactions have to be performed for each sample to ensure accuracy. Commonly, two or more segments of both HVI and HVII are sequenced in both directions and the $>$800 bases of combined sequence must be reviewed for each sample. Therefore, due to its speed of analysis and discrimination, the immobilized SSO probe strip typing system is ideal for preliminary analysis of samples. If additional sequence information is required, then sequence analysis can be performed on a limited number of samples that were not excluded. The casework example described here clearly demonstrates how these two methods can be used together to reduce casework turn-around time significantly. Also, due to the different detection limitations of the two approaches, the most complete and accurate information about a particular sample will be obtained when both methods are used. Sometimes due to limited material, PCR inhibitors, or DNA degradation the amount of PCR product generated is sufficient for only one method of analysis. In those laboratories with the capability to perform sequence analysis, the analyst should proceed directly to sequencing.

The most critical and challenging areas of mtDNA typing are the collection, extraction, and manipulation of the samples, not the development of methods for detecting sequence variation in mtDNA. The amount of mtDNA in a single cell is sufficient for detection in most PCR-based typing assays, particularly in the negative controls and in samples with little mtDNA. Therefore, additional precautions need to be taken during sample collection when possible, and laboratory space and equipment should be dedicated to mtDNA analysis. Also, analysts should be very experienced with PCR-based assays and perhaps should receive specialized training and proficiency testing for mtDNA typing, whether it is performed by using immobilized SSO probe strips or automated sequence analysis. While many samples can be analyzed at the same time by both immobilized SSO probe and sequencing methods, the extraction and amplification steps should be carried out on one sample at a time. For shed hair evidence, external contaminants such as blood, saliva, and semen should be removed prior to extraction (e.g., by sonicating hair in the presence of SDS). Similarly, contaminating material on bones and teeth can be removed by grinding prior to extraction or soaking in 10% bleach. Unfortunately, contaminating DNA cannot be removed from other types of evidence prior to extraction, and low levels of contamination from various unknown sources have, on occasion, been observed in negative controls (34). However, by running multiple controls with each set of extractions and amplifications, it usually is possible to determine if the contamination has compromised the data. Laboratories will need to develop additional DNA typing guidelines for interpreting and supporting mtDNA results containing a low amount of contamination.

Acknowledgments

We thank M. Grow for probe design and S. Walsh for designing the HVII primers and providing valuable input during the early stage of this project. The authors also thank S. Calloway for her thoughtful comments on the manuscript, U. Gyllensten for valuable discussions, and T. Marshak for creating the sequence database query program. We thank C. Mano, J. Allen, J. Griggs, and S. Calloway for their DNA sequencing contributions and M. Maramba for her strip typing contribution. Finally, we thank the oligonucleotide synthesis groups at RMS for making the SSO probes.

References

1. Wallace DC, Garrison K, Knowler WC. Dramatic founder effects in Amerindian mitochondrial DNAs. Am J Phys Anthropol 1985;68: 149–55.
2. Cann RL, Stoneking M, Wilson AC. Mitochondrial DNA and human evolution. Nature 1987;325:31–6.
3. Harihara S, Saitou N, Hirai M, Gojobori T, Park KS, Misawa S, et al. Mitochondrial DNA polymorphism among five Asian populations. Am J Hum Genet 1988;43:134–43.
4. Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC. Mitochondrial DNA sequences in single hairs from a southern African population. Proc Natl Acad Sci USA 1989;86(23):9350–4.
5. Horai S, Hayasaka K. Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. Am J Hum Genet 1990;46:828–42.
6. Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC. African populations and the evolution of human mitochondrial DNA. Science 1991;253:1503–7.
7. Ward RH, Frazier BL, Dew-Jager K, Pääbo S. Extensive mitochondrial

diversity within a single Amerindian tribe. Proc Natl Acad Sci USA 1991;88:8720–4.
8. Torroni A, Schurr TG, Yang C-C, Szathmary EJE, Williams RC, Schanfield MS, et al. Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. Genetics 1992;130:153–62.
9. Hsieh C-L, Sutton HE. Mitochondrial and nuclear variants in a U.S. Black population: origins of a hybrid population. Ann Hum Genet 1992;56:105–12.
10. Monsalve MV, Hagelberg E. Mitochondrial DNA polymorphisms in Carib people of Belize. Proc R Soc Lond B 1997;264:1217–24.
11. Comas D, Calafell F, Mateu E, Pérez-Lezaun A, Bosch E, Martínez-Arias R et al. Trading genes along the Silk Road: mtDNA sequences and the origin of Central Asian populations. Am J Hum Genet 1998; 63:1824–38.
12. Stoneking M, Hedgecock D, Higuchi R, Vigilant L, Erlich HA. Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. Am J Hum Genet 1991;48:370–82.
13. Melton T, Peterson R, Redd AJ, Saha N, Sofro AS, Martinson J, et al. Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. Am J Hum Genet 1995 Aug;57:403–14.
14. Melton T, Stoneking M. Extent of heterogeneity in mitochondrial DNA of ethnic Asian populations. J Forensic Sci 1996 Jul;41(4):591–602.
15. Melton T, Wilson M, Batzer M, Stoneking M. Extent of heterogeneity in mitochondrial DNA of European populations. J Forensic Sci 1997 May;42(3):437–46.
16. Melton T, Ginther C, Sensabaugh G, Soodyall H, Stoneking M. Extent of heterogeneity in mitochondrial DNA of sub-Saharan African populations. J Forensic Sci 1997 Jul;42(4):582–92.
17. Soodyall H, Jenkins T, Mukherjee A, Du Toit E, Roberts DF, Stoneking M. The founding mitochondrial DNA lineages of Tristan da Cunha islanders. Am J of Phys Anthropol 1997;104:157–66.
18. Wong L-JC, Senadheera D. Direct detection of multiple point mutations in mitochondrial DNA. Clin Chem 1997;43:1857–61.
19. Satoh M, Kuroiwa T. Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. Exp Cell Res 1991; 196:137–40.
20. Giles RE, Blanc H, Cann HM, Wallace DC. Maternal inheritance of human mitochondrial DNA. Proc Natl Acad Sci USA 1980;77:6715–9.
21. Anderson S, Bankier AT, Barrell BG, deBruijn MHL, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. Nature 1981;290:457–65.
22. Ikuta S, Takagi K, Wallace RB, Itakura K. Dissociation kinetics of 19 base paired oligonucleotide-DNA duplexes containing different single mismatched base pairs. Nuc Acids Res 1987;15:797–811.
23. Higuchi R, von Beroldingen CH, Sensabaugh GF, Erlich HA. DNA typing from single hairs. Nature 1988;332:543–6.
24. Allen M, Engström A-S, Meyers S, Handt O, Saldeen T, von Haeseler A, et al. Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: sensitivity and matching probabilities. J Forensic Sci 1998;43 (3):453–64.
25. Wilson MR, DiZinno JA, Polanksey D, Replogle J, Budowle B. Validation of mitochondrial DNA sequencing for forensic casework analysis. Int J Legal Medicine 1995;108:68–74.
26. Comas D, Reynolds R, Sajantila A. Analysis of mtDNA HVRII in several human populations using an immobilized SSO probe hybridization assay. Euro J Hum Genet 1999;7:459–68.
27. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 1989;123:585–95.
28. Bendall KE, Macaulay VA, Baker JR, Sykes BC. Heteroplasmic point mutations in the human mtDNA control region. Am J Hum Genet 1996;59:1276–87.
29. Jazin EE, Cavelier L, Eriksson I, Oreland L, Gyllensten U. Human brain contains high levels of heteroplasmy in the noncoding regions of mitochondrial DNA. Proc Natl Acad Sci USA 1996;93:12382–7.
30. MITOMAP: Human Mitochondrial Genome Database. 1998 Center for Molecular Medicine, Emory University, Atlanta, GA http://www.gen.emory.edu/mitomap.html.
31. Linch CA, Smith SL, Prahlow JA. Evaluation of the human hair root for DNA typing subsequent to microscopic comparison. J Forensic Sci 1998;43:305–14.
32. Wilson MR, Polanskey D, Replogle J, DiZinno JA, Budowle B. A family exhibiting heteroplasmy in the human mitochondrial DNA control region reveals both somatic mosaicism and pronounced segregation of mitotypes. Hum Genet 1997;100:167–71.
33. Calloway CD, Reynolds RL, Herrin, Jr GL, Anderson WW. The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. Am J Hum Genet 2000;66:1384–97.
34. Wilson MR, DiZinno JA, Polanskey D, Replogle J, Budowle B. Validation of mitochondrial DNA sequencing for forensic casework analysis. Int J Legal Med 1995;108:68–74.

Additional information and reprint requests:
Rebecca L. Reynolds, Ph.D.
Roche Molecular Systems
1145 Atlantic Avenue
Alameda, CA 94501-1145