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

Matthew N. Gabriel, Cassandra D. Calloway¹, Rebecca L. Reynolds¹, Dragan Primorac²

San Francisco Police Department, Criminalistics Laboratory, San Francisco; ¹Roche Molecular Systems, Department of Human Genetics, Alameda, CA, USA; and ²Split University Hospital and School of Medicine, Laboratory for Clinical and Forensic Genetics, Split, Croatia

Aim. A rapid analysis of mitochondrial DNA (mtDNA) sequences with an array of immobilized sequence-specific oligonucleotide (SSO) probes was tested on 18 skeletal elements recovered from mass graves in Croatia, which could not be genotyped with common forensic nuclear DNA systems (PM+DQA1 and short tandem repeat analysis).

Methods. We used duplex polymerase chain reaction (PCR) amplification of the mtDNA hypervariable regions I and II (HVI and HVII) (444 bp and 415 bp amplicons, respectively) and subsequent linear array typing, which targets six polymorphic regions and two additional sites within the human mtDNA HVI and HVII. The remaining amplified products were subjected to direct sequence analysis to obtain complete sequence information for the targeted HV regions.

Result. Duplex PCR amplification of the mtDNA HVI and HVII was successful in providing sufficient product for typing with the array of SSO probes in 14 out of the 18 sample extracts. We report here the sequence match of one set of remains with a panel of immobilized SSO probes, followed by direct sequence analysis. The corresponding mtDNA haplotype obtained for the bone sample and the putative maternal reference was unique in a database of 105 randomly selected Croatian individuals.

Conclusion. Mitochondrial DNA typing with an array of immobilized SSO probes can be a benefit to forensic DNA analysis of mass disaster remains and identity testing of single and mass graves.

Key words: Croatia; DNA, mitochondrial; forensic anthropology; forensic medicine; oligonucleotide probes

Analysis of human mitochondrial DNA (mtDNA) control region sequences is a useful tool for forensic identity testing on a range of samples (1-3) due to the high sequence variability (4), genome copy number per cell (5), and maternal inheritance (6). Sequence variation within the human mtDNA control region has been widely established as a valuable marker for individual identification (7-11). Polymerase chain reaction (PCR)-based techniques used to detect variation include direct DNA sequence analysis and screening with a panel of sequence-specific oligonucleotide (SSO) probes (3,9,10,12-20). These studies have established a high degree of polymorphism located primarily within two regions of the human mtDNA control region known as hypervariable regions I and II (HVI and HVII). The high mtDNA genome copy number per cell (∼500-1,000 copies) allows analysis of the most challenging forensic specimens, such as shed hairs from crime scenes and significantly aged remains (1-4,17,21-24). For this reason, mtDNA testing is often successful in cases where nuclear markers cannot be amplified. In addition, the maternal inheritance of mtDNA genomes allows a greatly expanded reference sample population for human identification efforts.

Typing of human mtDNA sequences by using SSO probes has been described elsewhere (12,17,25,30). Reynolds et al (17) reported the detection of sequence variation for 689 unrelated individuals, using a typing strip that contained an array of immobilized SSO probes specific for variants within five polymorphic regions of HVII. This method utilizes immobilized SSO probes for rapid sample screening in a reverse line blot format. From this study, the panel of 16 immobilized SSO probes for HVII was shown to provide a significant power of discrimination, ranging from ∼0.92-0.98, among African American, US Caucasian, US Hispanic, and Japanese populations (17). The original typing strip has since been expanded to include four polymorphic regions and one site within HVI as well as an additional HVII site for sequence variation detection (HVI A, C, D, E, and HVII A, B, C,
D, and positions 189 and 16093 relative to the Anderson reference sequence (12,26). Two versions of the expanded HVI/HVII typing strip were used in this analysis. The first version contains 27 possible immobilized SSO probe signals (HVI A, C and HVII A, B, C, D, and 189 and 16093), whereas the fully expanded array contains 31 possible probe signals (HVI A, C, D, E and HVII A, B, C, D, and 189 and 16093) for HVI and HVII. Additional probes are currently being investigated as potential candidates for sequence variation detection. Although there is a greater degree of polymorphism within HVI than HVII, the distribution of variable positions across HVII is less clustered than HVI. Therefore, fewer candidate HVII polymorphic regions are ideal for analysis by probe hybridization methods.

Despite the relatively high success rate obtained with nuclear PM+DQA1 and short tandem repeat analysis (~85%) on bone samples recovered from similar mass graves, the eighteen extracts tested here could not be amplified with nuclear markers (27). In this paper, we report a duplex PCR amplification strategy for full-length HVI and HVII amplicons of 444 bp and 415 bp, respectively, and subsequent typing of skeletal elements recovered from mass graves in Croatia (27,28).

**Material and Methods**

**DNA Extraction from Bone**

Samples of long bones were collected for DNA analysis at the time of autopsy. Bones were cleaned from remnant soft tissue and soil traces, rinsed several times in distilled water, and air-dried. The exterior was sanded to remove soil traces and the bone fragment was rinsed twice with 5% commercial bleach (Hygine Plus, Germany). The resultant 2-5 g bone fragment was rinsed twice with 5% commercial bleach, followed by deionized water and 80% ethanol washes. Samples were air-dried for 3-4 h, placed in steel-plated chambers, and followed by a single chloroform extraction. A 3 mL n-butanol extraction was performed twice with equal volumes, followed by a 12-minute activation step at 92 °C. Samples were subjected to 38 cycles at 92 °C for 30 s, 60 °C for 30 s, and 72 °C for 30 s, followed by a final 72 °C extension step for 10 min. Sequences for the biotinylated HVI and HVII primers are as follows: HVI: F13975 5’-XCTCCACATTAGACCCCA-3’ and R16418 5’-XATTTCACCAGCGATGGTGC-3’. HVII: F15 5’-XCATTTAACCACCTCAG-3’ and R429 5’-XCTGTTAAAG TGCATACCCCGA-3’ (where X denotes biotin). PCR products were analyzed on 1.5% agarose gels in 1X Tris-Borate-EDTA electrophoresis buffer (TBE) stained with 0.5 µg/ml ethidium bromide. Products were used directly for SSO typing and then purified with Millipore Ultrafree-MC® centrifugal filter devices (Millipore, Bedford, MA, USA) to eliminate excess primer prior to sequence analysis. For filtration, 25-40 µL of PCR product was added to 350 µL of Tris-EDTA (TE), pH 8.0, and centrifuged for 4 min at 12,000 rpm. After 20 µL of TE was added to the membrane, the recovery was transferred to a clean microcentrifuge tube for cycle sequencing.

**Immobilized SSO Probe Design**

A complete description of immobilized SSO probe design has been reported elsewhere (17), and those guidelines were followed for designing additional probes in the A, C, D, and E regions of HVII. In general, the ascending order of destabilizing mismatches is as follows: A-T,G-C << C-T,G-T,C-C,A-G,A-G-G,G,A-A,C-A << T-T,C-T,C-C. Immobilized SSO Probe Strip Typing

PCR product yield was compared to a low DNA mass ladder (Life Technologies, Gaithersburg, MD, USA) and roughly 20-150 ng was used for immobilized SSO typing. The typing protocol follows similar conditions as the AmpliType PM/PA+ DQA1 protocol (PE Biosystems) with the following modifications (17):

1. PCR product was prepared for hybridization by mixing 1:1 with AmpliCor® denaturant solution (1.6% NaOH (Roche Diagnostics, Branchburg, NJ, USA) and incubated at room temperature for 5 min.
2. The assay was performed at 55°C ± 1 °C.
3. Wash solution consists of 2X SSPE, 0.5% SDS.
4. Eight µL of enzyme conjugate (HRP-SA (PE Biosystems) is added per strip.
5. Four µL of 3% hydrogen peroxide per strip is added to the color development solution.
6. The volume of all typing solutions is 3 µL per well.

**Mitotype Designations**

Figure 1 shows the mtotype designations for several samples used in this study. Visible probe signals were given number designations for each probe region. For example, sample extract 52 provided visible signal for the following probes: HVI: A1, C1, and HVII - A1, B4, C2, D1, 189(A), and 16093(T). Therefore, the mtotype for this sample was designated “111421(A/T)”. Mitotype designations were specific to the array of probes that are used for a given analysis. Bone extracts were initially typed with an array that consists of HVI A and C, HVII A, B, C, D, and positions 189 and 16093 (30). The data obtained were sufficient to determine sample exclusions for all but one of the unknowns (data not shown). A subsequent expanded version of the array that includes probes for HVI D and E regions has been developed, and several samples in this study were successfully typed with this more informative array. This fully expanded version of the array has been optimized for use in the forensic DNA community and included in the LINEAR ARRAY™ mtDNA HVI/HVII Region-Sequence Typing Kit (Roche Applied Sciences, Indianapolis, IN, USA). The remaining samples were consumed during the initial analysis and therefore could not be typed with the expanded linear array.
There are four types of probe signal patterns that can occur at each polymorphic region. The most commonly observed strip typing result is a single visible probe within each of the regions. DNA sequences that are not complementary to any of the probes for a given region result in the absence of probe signal. A '0' designation is given to a region with no detectable signal. The '0' types do not necessarily have the same sequence because there are a number of variant sequences that will prevent hybridization to the probes for a given region. Further, a 'W' designation is given to signals that appear weaker than the probe intensities in other regions. This weak signal is caused by the presence of additional variants that act to partially destabilize the template-probe interaction. Lastly, multiple probe signals observed within a single polymorphic region suggest either the presence of mixed sequences due to a DNA contaminant (multiple DNA contributors) or to mtDNA heteroplasmy (presence of two detectable mtDNA sequences within an individual).

**Cycle Sequencing**

For cycle sequencing with BigDye™ Dye Terminator chemistry (PE Biosystems), 5-20 ng of the purified product were used. Cycle sequencing was performed in 20 μL volumes with 8.0 μL of BigDye™ Terminator mix and 1.6 μmol/L of the appropriate forward or reverse primer used for amplification. Cycle sequencing was performed as follows: 96 °C for 15 seconds, 50 °C for 5 seconds, and 60 °C for 4 minutes for a total of 25 cycles. One positive sequence control was performed for each group of samples with pGEM plasmid and M13 primer. Isopropanol precipitation of the product was performed by the addition of 80 μL of 75% isopropanol with mixing followed by incubation at room temperature for 45 minutes. Centrifugation was carried out at 3,500 rpm for 45 minutes at room temperature. Samples were then inverted and the isopropanol residue was removed by centrifugation. Samples were resuspended in 20 μL ABI PRISM template suppression reagent (PE Biosystems) or 5 μL 6× 50 mmol/L EDTA + deionized formamide (1:5) loading buffer for ABI PRISM™ 310 capillary electrophoresis injection or polyacrylamide gel loading, respectively, and denatured by heating at 92 °C for 3 minutes. For capillary electrophoresis, 10-30 second injections were performed with a 47 cm capillary and ABI PRISM POP-6™ polymer (PE Biosystems). For samples run on the 377 DNA Sequencer, 1.5 μL of the 6 μL resuspension volume was electrophoresed on a 4% polyacrylamide gel (19:1) prepared from Long Ranger® Singel® packs (FMC Bioproducts, Rockland, ME, USA). Following data collection, sequence data analysis was performed with Sequencher™ software (Gene Codes Corporation, Ann Arbor, MI, USA).

**Results**

Fourteen of 18 bone samples were successfully amplified using duplex PCR and typed with immobi-
Figure 3. Detection of mtDNA heteroplasmy for bone sample "S1". Shown above are partial HVII B and C region sequence information for bone sample S1. The mitotype designation for S1 is "1111111(A)(T)" as reported in Fig. 1. The presence of B1 and B3 signals for the HVII B region indicates a 152 C/T mixture, suggestive of mtDNA heteroplasmy. Panels A and B show the sequence data that confirm the suspected heteroplasmy at position 152 – a majority 'T' peak is observed with underlying 'C' at this position. Sequence data were obtained from independent amplifications.

Discussion

SSO typing of sequence polymorphisms for human identification efforts can provide significant benefit in many forensic DNA analyses. Reassociation of mass disaster remains, identity testing for remains recovered from single and mass graves, and screening of large numbers of crime scene samples for inclusion or exclusion of a potential suspect(s) are instances in which SSO typing can aid forensic identity testing. Further, co-amplification of the HVI and HVII regions and typing with the immobilized SSO probe strip assay expands the number of samples that can be processed in a given period of time. For this study, 18 bone sample extracts (plus controls) and a separate set of 4 maternal references were amplified and typed in a single day. A preliminary sequence match was made, and only one bone sample product and the putative match were subjected to direct sequence analysis using a total of 8 sequencing reactions (forward and reverse sequencing directions for each HV region). The corresponding mtDNA haplotype obtained for the bone sample and the putative maternal reference was unique in a database of 105 randomly selected Croatian individuals (30). No additional amplification reactions were required because the duplex PCR products served as sequencing templates for all of the reactions. If the same set of samples had been processed with single target PCR amplification followed by direct sequencing methods, a minimum of 44 amplifications (plus controls) and 88 separate sequencing reactions would have been required for comparisons. The SSO screening approach and sequence analysis is therefore beneficial in directing an analyst toward those specimens that possess the greatest evidentiary value for a given case.

In this study, seven samples yielded ambiguous typing results. Four of the samples could not be ampli-
fied with the PCR primers for ~400 bp products. These samples may lack sufficient DNA for amplification, consist of DNA that is too degraded for analysis, or contain enzymatic inhibitors that prevent PCR amplification. Environmental factors including soil acidity and composition, heat, and humidity contribute to the fragmentation of DNA molecules, and therefore the extent of degradation can vary greatly depending on the burial location. If one or more of these four samples contain degraded DNA, then amplification with primers that target smaller products might yield sufficient PCR product for analysis. The remaining three samples contained DNA sequences from more than one source. This result was not surprising as mtDNA laboratories occasionally encounter sporadic, low-level contamination from undetermined sources (4,21). These mixtures presumably result from the co-amplification of authentic sample template and human mtDNA contaminants present either in the reagents or intrinsic to the sample. In the three samples that contain multiple sequences, the authentic DNA template may be highly degraded, whereas the contaminating DNA is likely to be intact. In the most extreme cases of degradation, primers that target short products (~100-150 bp) are effective for successful amplification of authentic DNA sequences even in the presence of low-level contaminants (21-23). Two separate duplex amplifications that target smaller ~270 bp products were used to amplify the samples with mixed sequences and those that did not amplify with the larger HV region products, and similar results were obtained (unpublished observation). The extracts were consumed during testing and therefore attempts to target ~100-150 bp amplicons could not be performed.

The full-length HVI and HVII duplex PCR amplification described here is useful on a range of forensic specimens, including shaft portions of single hairs, blood, and body tissues, and is sensitive to ~10 pg of high quality total genomic DNA. Further optimization has resulted in a duplex PCR assay sensitive to <1 pg of high quality total genomic DNA. The immobilized SSO probe strips have been shown to be more consistent and sensitive than direct sequence analysis in samples containing known ratios of two sequences (17) and in heteroplasmic tissue samples (25). In this study, bone sample S1 contained two roughly equal probe signal intensities in the HVI B region, suggesting the presence of mtDNA heteroplasm. Heteroplasmy at position 152 was confirmed with direct sequence analysis, which showed the presence of a major 'C' peak with underlying 'T'. No additional positions suggested the presence of multiple sequences, indicating that the mixture at 152 was likely due to mtDNA heteroplasm and not contamination from a second source. Heteroplasmy at position 152 has been previously observed in several sample types other than bone (17,25), which suggests this site is a potential hotspot for heteroplasmy. A high frequency of heteroplasmy has been reported at additional sites, including 16093 (31) and 189 (25). Although heteroplasmy may occur at some positions more frequently than others, it can potentially occur at any site. As with mtDNA sequence analysis interpretation guidelines, careful consideration should be given to instances of heteroplasmy when determining interpretation guidelines and evaluating forensic cases in which heteroplasmy may have occurred. Similar guidelines can be developed for use with the LINEAR ARRAY™ mtDNA HVI/HVII Region-Sequence Typing Kit (Roche Applied Sciences) to determine if a sample can be excluded based on mitotyping results alone or should be submitted for subsequent analysis. The authors suggest that analysts base their considerations for sequence analysis on the number of nucleotide differences as defined by the corresponding probe signals obtained with the LINEAR ARRAY™ Typing Kit. We are currently developing a software program to convert mitotypes to their respective nucleotide base designations to aid in making these types of analytical determinations as well as for querying the mtDNA Combined DNA Index System (CODIS) database. Once a set of suitable interpretation criteria have been established by a given laboratory, we suggest that samples identified as "inconclusive" or "cannot be excluded" because of shared mitotyping DNA with the linear array be submitted for sequence analysis. Although the discrimination power of the mtDNA linear array is considerable, the array targets a subset of variable positions within the hypervariable regions and, therefore, additional variation may be observed with sequence analysis.

The immobilized SSO probe screening strategy for forensic mtDNA typing has the potential to reduce the total sample number for a given case to a smaller subset of samples that provide evidentiary value. This advantage may be attractive to forensic identity testing and criminal investigation efforts that otherwise cannot afford to perform unnecessary sequence analysis due to time and budget constraints. A summary of forensic casework analyses that benefited from this screening strategy by excluding numerous crime scene samples will be described elsewhere (manuscript in preparation). Further optimization of the linear array and duplex PCR assay described here has been performed to provide the forensic DNA community with an even more robust mtDNA typing kit. A final version of the LINEAR ARRAY™ mtDNA HVI/HVII Region-Sequence Typing Kit (Roche Applied Sciences) has been tested by multiple forensic laboratories, and the results of these studies as well as in-house developmental validation studies, recommended assay conditions, and data interpretation for the typing kit will be reported elsewhere.

Acknowledgment

The authors thank Henry Erlich, Karen Walker, Michael Grow, Natasha Stankiewicz, and Sahar Naderi for scientific and intellectual contributions and laboratory assistance, and Irena Drmić for sample collection. This project was supported by National Institute of Justice grant No. 96-IJ-CX-0028.

References
