# Typing the 1.1 kb Control Region of Human Mitochondrial DNA in Japanese Individuals 


#### Abstract

This study presents a reliable method that uses high-fidelity long-range PCR and optimized primers to assess polymorphism and to genotype human mitochondrial DNA (mtDNA). This method was used to analyze polymorphic sites in the human mtDNA control region, including hypervariable regions I, II, and III (HVI, HVII, and HVIII), from 124 unrelated Japanese individuals. In HVI, HVII, and HVIII, 80, 37, and 14 polymorphic sites were identified, respectively, excluding those in the homopolymeric cytosine stretch (C-stretch) regions. The region between HVI and HVII also contained 15 polymorphic sites. On the other hand, C-stretch length heteroplasmy in HVI or HVII was observed in 66 of 124 Japanese individuals (53\%), which is much higher than in Caucasian populations. The variants in the C-stretch regions were characterized by counting the number of heteroplasmic peaks split from the single peak in homoplasmic sequences (i.e., ${ }^{16244} \mathrm{G}$ and ${ }^{16255} \mathrm{G}$ in HVI and ${ }^{285} \mathrm{G}$ in HVII). Including the C-stretch length heteroplasmy, the 124 Japanese mtDNA samples were classified into 116 distinct haplotypes. The random match probability and the genetic diversity were estimated to be $0.95 \%$ and 0.998581 , respectively, indicating that the method presented here has higher discrimination than the conventional method for mtDNA typing using HVI and HVII. [Correction added after publication 30 January 2007: in the preceding sentence random match probability and genetic diversity estimates were corrected from 0.95 and $0.998581 \%$, respectively, to $0.95 \%$ and 0.998581 , respectively.] The haplogroups and their frequencies observed in this study (i.e., D4; 13.7\%, M7a1; 11.3\%, D4a; 9.7\% and M7b2; $8.9 \%$ ) were similar to those observed in other studies of Japanese mtDNA polymorphism. The method described here is suitable for forensic applications, as shown by successful analysis of tissues from highly putrefied remains of an infant, which allowed maternal relationship to be determined via mtDNA haplotyping.


KEYWORDS: forensic science, mitochondrial DNA, sequence polymorphism, Japanese population, hypervariable regions, maternal identification, haplogroup

Polymorphic sites in human nuclear DNA are routinely used in forensic sciences to uniquely match biological samples with the individual from whom they came (see reviews (1,2)). Using a PCR-based technology, several types of polymorphism are routinely identified in nuclear DNA including base substitutions, insertions/deletions, or variation in length/number of tandem repeats. However, when the amount of nuclear DNA is limiting or when the sample is highly degraded, it can be difficult to obtain data that are sufficiently reliable for forensic applications. In such cases, polymorphism for genotyping can be obtained by sequencing mitochondrial DNA (mtDNA), which is present in hundreds to thousands of copies per cell and that is more stable than nuclear DNA. mtDNA, unlike nuclear DNA, is maternally inherited, which makes it possible to identify maternal relationships even for individuals who are separated by more than one generation. In addition, there is no evidence that mtDNA undergoes homologous recombination, so mtDNA is considered a haploid genome and its sequence can be determined directly by PCR. Thus, mtDNA typing is considered a very useful tool for forensic identification (see reviews (3-5)) despite the fact that the mode of inheritance and lack of recombination results in reduced power of discrimination relative to nuclear DNA genotyping. This drawback can be overcome by constructing large databases of mtDNA haplotypes.

[^0]Human mtDNA has three restricted polymorphic segments within the control region designated as hypervariable regions I (HVI; $342 \mathrm{bp}, 16,024-16,365$ ), II (HVII; $268 \mathrm{bp}, 73-340$ ), and III (HVIII; $137 \mathrm{bp}, 438-574)(6,7)$. mtDNA typing often focuses on polymorphism in HVI and HVII (8-15). Lutz et al. (16) reported that about one-fifth of all mtDNA genomes that are identical in HVI and HVII had nonidentical sequences in HVIII. Recently, studies on mtDNA typing including hypervariable regions HVI, II, and III have increased (7,16-21) and some mutations have been reported in the control region and the coding region of human mtDNA (see reviews $(22,23)$ ). Therefore, a database of the mtDNA sequences covering the entire control region might be useful.

This study presents a reliable PCR-based method for identifying polymorphism in the 1.1 kb control region of human mtDNA. The method uses long-range PCR with engineered high-fidelity Taq polymerase (24) and optimized primers to analyze mtDNA from 124 unrelated Japanese individuals. The method described here is suitable for forensic applications, as shown by successful mtDNA haplotyping of tissues from highly putrefied remains of an infant.

## Materials and Methods

The study was approved by the Ethics Commission at the Faculty of Medicine, University of Yamanashi, and the volunteers were informed about the aim of the study and written consent was obtained.

## Sample Preparation

During DNA manipulations, all workers wore a cap, mask, and gloves at all times to reduce the risk of contamination.

For population studies: Venous blood samples were collected with informed consent from 124 unrelated Japanese individuals. DNA was extracted from $40 \mu \mathrm{~L}$ blood by the proteinase $\mathrm{K} /$ sodium
dodecyl sulfate (SDS) method. DNA was extracted with phenol/ chloroform as described previously (25) and dissolved in $20 \mu \mathrm{~L}$ 10 mM Tris-HCl, 1 mM EDTA pH 8.0 (TE).

For forensic case study: Three tissue specimens were obtained during autopsy of highly putrefied remains of an infant; specimens were from the umbilical cord, skin of the nuchal region, and intercostal muscle. A blood sample on a piece of gauze was obtained from the putative mother. All samples were stored at $-20^{\circ} \mathrm{C}$ before analysis. Tissue samples from the infant (50100 mg each) and the blood sample from the putative mother $(1 \times 1 \mathrm{~cm})$ were cut into $c .2 \mathrm{~mm}$ pieces. Each sample was placed in a 1.5 mL plastic centrifuge tube. One milliliter extraction buffer ( 10 mM Tris- $\mathrm{HCl}, 100 \mathrm{mM} \mathrm{NaCl}, 10 \mathrm{mM}$ EDTA pH 8.0), $100 \mu \mathrm{~L}$ $10 \% \mathrm{SDS}$, and $50 \mu \mathrm{~L} 20 \mathrm{mg} / \mathrm{mL}$ proteinase K were added to the samples. Forty microlitres of 1 M dithiothreitol was also added to the tissue samples. The mixture was incubated overnight at $56^{\circ} \mathrm{C}$ on a rotary mixer and divided into two aliquots. One volume of phenol equilibrated with TE was added to 1 aliquot and mixed by slow rotation for 2 h at room temperature. The mixture was centrifuged at $15,000 \mathrm{r} . \mathrm{p} . \mathrm{m}$. for 20 min at room temperature. The aqueous phase was transferred to a new 1.5 mL tube, and extracted once with phenol, twice with phenol/chloroform/isoamylalcohol (25:24:1), and once with chloroform/isoamylalcohol (24:1). DNA in the acqueous phase was precipitated with 0.1 volume of 3 M sodium acetate ( pH 5.2 ) and 2.5 volume of ethanol at $-20^{\circ} \mathrm{C}$ overnight. The DNA was collected by centrifugation at $15,000 \mathrm{r} . \mathrm{p} . \mathrm{m}$. for 20 min at $4^{\circ} \mathrm{C}$, washed with $70 \%$ ethanol, and then dissolved in $30 \mu \mathrm{~L}$ TE. The DNA yield was estimated from absorbance at 260 nm . DNA integrity was evaluated by agarose gel electrophoresis using $1 \mu \mathrm{~L}$ extract.

## PCR Amplification and Purification of PCR Products

A first primer set, H16044 (16,025-16,044, 5'-GAG-GGGTGGCTTTGGAGTTG-3') and L544 (563-544, 5'-GCCTAAATAGCCCACACGTTCC-3'), was used to amplify $c$. 1.1 kb in length containing the almost entire human mtDNA control region. The primer set was designed to avoid co-amplification of nuclear-embedded mtDNA (NUMT) sequences. The absence of potential primer-binding sites in NUMT DNA was confirmed by conducting a BLAST search using the primer sequence as a query against human genome database. PCR was performed in a TaKaRa thermal cycler 480 (Takara-Bio, Shiga, Japan). The reactions were carried out in $50 \mu \mathrm{~L}$ reaction mixture containing $5 \mu \mathrm{~L}$ 10x Ex Taq buffer (Takara-Bio), $2 \mathrm{mM} \mathrm{MgCl} 2,200 \mu \mathrm{M}$ dNTPs, $0.2 \mu \mathrm{M} \mathrm{H} 16044$ primer, $0.2 \mu \mathrm{M}$ L544 primer, and 1.25 units of TaKaRa Ex Taq ${ }^{\text {TM }}$ (Takara-Bio) composed with engineered highfidelity Taq polymerase and polymerase having proof-reading activity. The DNA template was 20 or 40 ng extracted DNA. The PCR consisted of initial 2 min denaturation at $96^{\circ} \mathrm{C}, 35$ cycles of 1 min denaturation at $94^{\circ} \mathrm{C}, 1 \mathrm{~min}$ annealing at $55^{\circ} \mathrm{C}$, a 3-min extension $72^{\circ} \mathrm{C}$, and a final $10-\mathrm{min}$ extension at $72^{\circ} \mathrm{C}$.

For highly putrefied infant tissue samples in forensic case study, 1 or $2 \mu \mathrm{~L}$ purified PCR products were subjected to a second round of PCR amplification under the same conditions.

## DNA Sequencing

The PCR products were separated on a $1 \%$ agarose gel, purified using the QIAquick Gel Extraction Kit (QIAGEN, Hilden, Germany), and used for DNA sequencing. Sequencing reactions were carried out with the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit Version 2.0 (Applied Biosystems,

Foster City, CA) using H16044, L20 (39-20), H16547 (1652616547), and L544 as sequencing primers. PCR products were sequenced on both strands by the ABI PRISM 310 Genetic Analyzer (Applied Biosystems). In this study, nucleotides in the mtDNA control region were numbered according to the revised version of Cambridge Reference Sequence (CRS) $(26,27)$.

Samples with poor resolution downstream of C-stretches were re-sequenced using L16245 (16264-16245) or H277 (258-277) primers located closer to the C-stretches and the ABI PRISM dGTP BigDye Terminator Cycle Sequencing FS Ready Reaction Kit (Applied Biosystems). In this case, the samples were amplified under the same conditions described above, except that TaKaRa LA Taq ${ }^{\text {TM }}$ and GC buffer II (Takara-Bio) were used. Reaction conditions were optimized to avoid the compression caused by repeated $\mathrm{G} / \mathrm{C}$ base pairs, so that the sequence of the C -stretch region could be determined precisely. Under the chosen conditions, heteroplasmic peaks were well resolved and could be readily counted. Polymorphisms at 4833, 4883, 5108, and 5178 were identified by nucleotide sequencing of PCR products amplified with primers, H4734 (4712-4734) and L5336 (5355-5336); this analysis classified some samples into groups G2a or D.

## Statistical Analysis

The probability of randomly selecting two samples with the same mitochondrial profile (random match probability) was calculated using the equation $p=\Sigma x^{2}$, where $x$ is the frequency of each mtDNA haplotype (28). Genetic diversity was estimated using the algorithm $\mathrm{h}=n\left(1-\Sigma x^{2}\right) /(n-1)$, where $n$ is the sample size and $x$ is the frequency of each mtDNA haplotype (29).

## Haplogroup Classification

mtDNAs were classified into haplogroups primarily according to the criteria proposed by Allard et al. (15) and secondarily according to criteria presented by Yao et al. (12), Kivisild et al. (30), and Kong et al. (31).

## Results and Discussion

## PCR Amplification of the 1.1 kb mtDNA Control Region

PCR primers (H16044 and L544) were optimized for amplification of the 1108 bp control region (16025-563) of human mtDNA without co-amplification of NUMTs (32). With these primers, the mtDNA control region was reliably amplified without by-products yielding sufficient PCR product for direct DNA sequence analysis. Then, DNA samples from 124 unrelated Japanese individuals were amplified and sequenced.

## Mutations in the 1.1 kb mtDNA Control Region

DNA sequences (16045-543) from 124 individuals were compared with each other and with the revised version of the Cambridge Reference Sequence (CRS; 26, 27). DNA sequencing revealed 148 polymorphic sites in the 124 samples (Table 1); 80 polymorphic sites were in HVI, 15 sites were in the region between HVI and HVII, 37 sites were in HVII, two sites were in the region between HVII and HVIII, and 14 sites were in HVIII. The fact that the region between HVI and HVII has as many polymorphic sites as HVIII suggests that it is worthwhile to analyze the control region, including non-HV sequences. Three single base heteroplasmic variants, T/C at 16362, A/G at 16399 and G/A at 185, were also found in three individuals among 124 samples (Table 1).
TABLE 1-Haplogroup and sequence variation in mtDNA control region of 124 Japanese individuals.

TABLE 1-Continued

TABLE 1—Continued


[^1]

FIG. 1-(a) Electropherograms of the sequences for homoplasmic mtDNA. The forward sequences were shown in HVI and the reverse sequences were shown in HVII. Arrowheads indicate peaks split from ${ }^{16193} \mathrm{C}$ in HVI and ${ }^{303} G$ in HVII. Arrows indicate peaks split from ${ }^{16244} G$ and ${ }^{16255} G$ in HVI and ${ }^{285} G$ in HVII. (b) and (c) Electropherogram of the sequences for heteroplasmic mtDNA. Arrowheads indicate the out-of-phase C peaks in HVI and Gs in HVII. Arrows indicate the peak splitting due to the heteroplasmy.

## C-Stretch Sequence Variants

The presence of C-stretches in HVI and HVII is well documented ( 33,34 ). In this study, 79 of 124 samples $(63.7 \%)$ had a homoplasmic C-stretch in HVI, and 75 samples ( $60.5 \%$ ) had a homoplasmic C-stretch in HVII. There were 58 samples (46.8\%) with homoplasmic C-stretches in HVI and HVII. Four and three sequence variants were identified in HVI and HVII, respectively. In HVI, ${ }^{16180} \mathrm{~A} 4 / \mathrm{C} 5 / \mathrm{T} / \mathrm{C} 4 / \mathrm{AT}^{16195}$ was the most common variant (57.3\%). In HVII, the variant ${ }^{303} \mathrm{C} 7 / \mathrm{T} / \mathrm{C} 6^{315.1}$ was present in 41 samples ( $33.1 \%$ ) and ${ }^{303} \mathrm{C} 8 / \mathrm{T} / \mathrm{C} 6^{315.1}$ was present in 32 samples $(25.8 \%)$. Similar frequencies were observed in other studies of mtDNA variation in Japanese populations $(9,11,13,14)$.

## C-Stretch Heteroplasmy

DNA samples with C-stretch heteroplasmy often show loss of resolution (i.e., characteristic out-of-phase sequence(s)) downstream of the C-stretch. In this dataset, a total of 66 samples ( $53.2 \%$ ) had loss of resolution after the C-stretch in HVI or HVII. In HVI, 45 samples ( $36.3 \%$ ) had loss of resolution downstream of C-stretch 16184-16193; this sequence included a T-C transition at 16189 (33), with or without an A-C transversion at 16,183 or two A to C transversions at 18182 and 16183 . Another 49 samples ( $39.5 \%$ ) had loss of resolution after C-stretch 303-309 in HVII with one to four C insertions.
When mtDNA polymorphism is used for forensic identification, C-stretch variation and heteroplasmy are not usually considered (35). However, because a large fraction of the 124 samples examined in this study have C-stretch length heteroplasmy compared with that in Caucasian populations ( $15 \%$ in HVI and $29 \%$ in HVII )
$(33,10)$, the following method was used to classify these variants. For G residues that gave a single peak in homoplasmic samples (i.e., ${ }^{16244} \mathrm{G}$ and ${ }^{16255} \mathrm{G}$ in HVI and ${ }^{285} \mathrm{G}$ in HVII; Fig. 1) and that appeared as split peaks in heteroplasmic samples, the number of split peaks were counted and assumed to represent the number of C-stretch variants. Peak height was neglected in this analysis because it does not correlate with abundance of a particular DNA species. The results are shown in Tables 1 and 2. Nine types of heteroplasmy were observed in HVI, with the most common type ( $13.7 \%$ ) being ${ }^{16180} \mathrm{~A} 3 / \mathrm{C} 10,11,12,13 / \mathrm{AT}^{16195}$; this sample has a mixed population including ${ }^{16180} \mathrm{~A} 3 / \mathrm{C} 10 / \mathrm{AT}^{16195},{ }^{16180} \mathrm{~A} 3 / \mathrm{C} 11 /$ $\mathrm{AT}^{16195},{ }^{16180} \mathrm{~A} 3 / \mathrm{C} 12 / \mathrm{AT}^{16195}$, and ${ }^{16180} \mathrm{~A} 3 / \mathrm{C}_{13} / \mathrm{AT}^{16195}$. Five types of heterplasmy were observed in HVII, with the most common two types being ${ }^{303} \mathrm{C} 8,9,10 / \mathrm{T} / \mathrm{C}^{315.1}$ and ${ }^{303} \mathrm{C} 8,9 / \mathrm{T} / \mathrm{C} 6^{315.1}$ ( $19.4 \%$ and $16.1 \%$, respectively). C-stretch heteroplasmy has been studied extensively, but C-stretch sequence variants are excluded from database searches. Thus, it is difficult to assess the variation observed in this dataset in relation to other datasets. However, some comparisons are possible and the results suggest that the observations are consistent between datasets. For example, one can compare the frequency of one type of heteroplasmic sample in our dataset with the sum of component frequencies in other datasets from Japanese populations (i.e., this dataset indicates a frequency of $16.1 \%$ for ${ }^{16180} \mathrm{~A} 3 / \mathrm{C} 9,10,11,12,13 / \mathrm{AT}^{16195},{ }^{16180} \mathrm{~A} 3 /$ C $10,11,12,13 / \mathrm{AT}^{16195}$, and ${ }^{16180} \mathrm{~A} 3 / \mathrm{C} 10,11,12,13,14 / \mathrm{AT}^{16195}$, while Horai and Hayasaka (36) report a frequency of $17.9 \%$ for ${ }^{16180} \mathrm{~A} 3 / \mathrm{C} 10 / \mathrm{AT}^{16195}, \quad{ }^{16180} \mathrm{~A} 3 / \mathrm{C} 12 / \mathrm{AT}^{16195}, \quad{ }^{16180} \mathrm{~A} 3 / \mathrm{C} 13 /$ $\mathrm{AT}^{16195}$, and ${ }^{16180} \mathrm{~A} 3 / \mathrm{C} 14 / \mathrm{AT}^{16195}$ ).

The analysis of C-stretch length heteroplasmy is potentially useful for forensic identification. The variability in C -stretch length heteroplasmy in different tissues from the same individu-

TABLE2A-Sequence variations in C-stretch from positions 16180 to 16195 in HVI.

|  | Sequence variations |  | $n=124$ |  | Heteroplasmy |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Type | CRS: AAAACCCCCTCCCCAT | A4/C5/T/C4/AT | $n$ | \% |  |  |  |
| 1 | AAAACCCCCTCCCCAT | A4/C5/T/C4/AT | 71 | 57.3 |  |  |  |
| 2 | AAAATCCCCTCCCCAT | A4/T/C4/T/C4/AT | 2 | 1.6 |  |  |  |
| 3 | AAAACCCTCTCCCCAT | A4/C3/TCT/C4/AT | 5 | 4.0 |  |  |  |
| 4 | AAAACCCCCCCCCAT | A4/C9/AT |  |  |  |  |  |
| 5 | AAAACCCCCCCCCCAT | A4/C10/AT |  |  | 8 | 3 |  |
| 6 | AAAACCCCCCCCCCCAT | A4/C11/AT |  |  | 6.5\% | 2.4\% |  |
| 7 | AAAACCCCCCCCCCCCAT | A4/C12/AT |  |  |  |  | 0.8\% |
| 8 | AAAACCСССССССССССАТ | A4/C13/AT |  |  |  |  |  |
| 9 | AAAACCCCCCCCCGC | A4/C9/GC |  |  | 2 |  |  |
| 10 | AAAACCCCCCCCCCGC | A4/C10/GC |  |  | 1.6\% |  |  |
| 11 | AAAACCCCCCCCCCCGC | A4/C11/GC |  |  |  |  |  |
| 12 | AAACCCCTCCCCCAT | A3/C4/T/C5/AT | 1 | 0.8 |  |  |  |
| 13 | AAACCCCCCCCCAT | A3/C9/AT |  |  |  |  |  |
| 14 | AAACCCCCCCCCCAT | A3/C10/AT |  |  |  |  |  |
| 15 | AAAСССССССССССАТ | A3/C11/AT |  |  | 1.6\% |  |  |
| 16 | AAACCCCCCCCCCCCAT | A3/C12/AT |  |  |  | 13.7\% | 0.8\% |
| 17 | АААСССССССССССССАТ | A3/C13/AT |  |  |  |  |  |
| 18 | AAACCCCCCCCCCCCCCAT | A3/C14/AT |  |  |  |  |  |
| 19 | AAССССССССССАТ | A2/C10/AT |  |  |  |  |  |
| 20 | AACCCCCCCCCCCAT | A2/C11/AT |  |  |  |  |  |
| 21 | AAССССССССССССАТ | A2/C12/AT |  |  | 7.3\% |  |  |
| 22 | AAСССССССССССССАТ | A2/C13/AT |  |  |  | 1.6\% |  |
| 23 | ААССССССССССССССАТ | A2/C14/AT |  |  |  |  |  |
| 24 | AACCCCCCCCCCCCCCCAT | A2/C15/AT |  |  |  |  |  |
| Total |  |  | 124 | 100.0 |  |  |  |

Heteroplasmic sequences are indicated by shadowed regions: components in each heteroplasmic variant are indicated by a vertical line.
al and in tissues that contain clonal populations of cells (i.e., hair) is well understood $(37,38)$. Then, C-stretch length heteroplasmy should only be compared in blood samples or samples from the same tissue for forensic applications.

## Haplotypes and Genetic Diversity

The sequences in this dataset were classified into 116 distinct haplotypes, which occurred with the following distribution: one haplotype was present in four individuals, another haplotype was present in three individuals, three haplotypes were present in two individuals, and 111 haplotypes were unique to single individuals. Statistical analysis estimated that two randomly selected individuals from this population have a probability of $0.95 \%$ of having identical mtDNA; in addition, the genetic diversity of the 1.1 kb mtDNA control region in this population was estimated to be 0.998581 (Table 3). The random match probability and the genetic diversity were also calculated for subregions within the mtDNA control region (Table 3). This analysis shows that fewer haplotypes (112) were identified, random match probability was slightly higher $(1.00 \%)$, and the genetic diversity was slightly lower ( 0.998065 ) when the calculation is based only on polymorphism
in HVI and HVII. Thus, the method presented here has higher discrimination power than the conventional method for analyzing variation in human mtDNA. Furthermore, genetic diversity was the same for this dataset and for a dataset analyzed by Vanecek et al. (20), which includes the DNA sequence of the D-loop region in 93 Czech Caucasians, but the random match probability was lower in the Japanese population analyzed here than for the Czech population. This difference is likely due to the larger number of DNA samples analyzed from the Japanese population and because Cstretch heteroplasmy was analyzed in the Japanese population.

## Haplogroups in 124 Samples

To assess consistency between the results of this and other studies of mtDNA variation, haplogroups and their frequencies were compared from these studies. This comparative analysis was based on haplogroups defined by variation in the 1.1 kb control region. Three samples could not be classified from these data alone, so the region between 4712 and 5355 was amplified and polymorphism was assessed at four additional sites: ${ }^{4833} \mathrm{G},{ }^{4883} \mathrm{~T}$, ${ }^{5108} \mathrm{C}$, and ${ }^{5178} \mathrm{~A}$. The results are shown in Table 1.

TABLE 2B—Sequence variations in C-stretch from positions 303 to 315 in HVII.

| Type | Sequence variations |  | $n=124$ |  | Heteroplasmy |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | CRS: CCCCCCCTCCCCC | C7/T/C5 | $n$ | \% |  |  |  |  |  |
| 1 | СССССССТСССССС | C7/T/C6 | 41 | 33.1 | 1 |  |  |  |  |
| 2 | ССССССССТСССССС | C8/T/C6 | 32 | 25.8 | 0.8\% | 20 | 24 | 3 |  |
| 3 | СССССССССТСССССС | C9/T/C6 |  |  |  | 16.1\% | 19.4\% | 2.4\% | 1 |
| 4 | ССССССССССТСССССС | C10/T/C6 |  |  |  |  |  |  | 0.8\% |
| 5 | СССССССССССТСССССС | C11/T/C6 |  |  |  |  |  |  |  |
| 6 | CCCCCCTCCCCCC | C6/T/C6 | 2 | 1.6 |  |  |  |  |  |
| Total |  |  | 124 | 100.0 |  |  |  |  |  |

[^2]TABLE 3-Genetic diversity and random match probability.

|  | HVI | HVII | HVIII | HVI+HVII | HVI+HVII + HVIII |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Genetic diversity | 0.985514 | 0.981110 | 0.726383 | 0.998065 | All |
| Random match probability | $2.24 \%$ | $2.68 \%$ | $27.95 \%$ | 0.998194 | 0.998581 |

Thirty distinct lineages were identified in the dataset from 124 Japanese individuals; 10 of these lineages were represented by one DNA sample. Four major macrohaplogroups were described in Japanese population previously $(14,15)$, and the frequency of these haplogroups was the same in this dataset as previously reported (D; $33.0 \%, \mathrm{M} ; 25.8 \%, B ; 13.6 \%$, and A; 9.6\%). Subhaplogroups D4 and D4a were the most common ( $13.7 \%$ and $9.7 \%$, respectively). Subhaplogroups M7a1 and M7b2 were also common ( $11.3 \%$ and $8.9 \%$, respectively). Macrohaplogroups F ( $5.6 \%$ ), $\mathrm{G}(5.6 \%)$, and $\mathrm{N}(4.8 \%)$ were less common in this dataset, but the most common subhaplogroups in F and N were F 1 b ( $4.8 \%$ ) and N9a (4.8\%) (Table 1). In summary, the dataset presented here, based on DNA samples from 124 unrelated Japanese individuals, includes similar haplogroup distribution and frequency as reported in other studies of mtDNA polymorphism in Japanese populations $(14,15)$.

## Case Study

In June 2001, the corpse of a newborn infant was found wrapped in a blanket inside a plastic bag in a piece of luggage. The bag in which the infant was hidden was left in a car parked at a shopping center. It was estimated that the infant corpse was found 3 weeks after it was left in the car. The corpse was highly putrefied but autopsy findings showed that the infant was born at full-term. The postmortem interval was thought to be 3-5 months. The woman who used the car was suspected to be the mother of the infant. mtDNA typing was performed using the method presented in this study to confirm maternal relationship between the suspect and deceased infant.

DNA was extracted from the umbilical cord, the skin of the nuchal region, and intercostal muscles of the infant corpse. The DNA yield varied from 14.4 to $42.2 \mathrm{ng} / \mathrm{mg}$ of muscle; values reported from fresh muscle were $457 \pm 190 \mathrm{ng} / \mathrm{mg}$ (39) or $607.9 \pm 310.5 \mathrm{ng} / \mathrm{mg}(40)$, suggesting that the DNA samples were highly degraded. Agarose gel electrophoresis confirmed that the DNA was degraded (Fig. 2a).

PCR amplification and mtDNA typing according to the method described here were carried out on blood from the putative mother


FIG. 2-(a) Electrophoresis of DNA extracted from the infant tissue on a $1 \%$ agarose gel. $M$, molecular weight marker (kb); lane 1, muscle; lane 2, umbilical cord; lane 3, skin; lane 4, blood from putative mother. (b) Electrophoresis of PCR products using DNA extracted from the infant tissues as template on a 1\% agarose gel. M, molecular weight marker (kb); lane 1, muscle; lane 2, skin; lane 3, umbilical cord; lane 4, blood from putative mother; $N$, negative control ( $P C R$ products amplified without template $D N A$ ).
and tissue samples from the infant corpse (Fig. 2b). DNA sequence was determined for both DNA strands, according to Guidelines for mtDNA typing (35). The results showed that the mtDNA of the infant and putative mother were identical (data not shown). The maximum probability that two unrelated individuals would have this mtDNA type is $1 / 124 \times 100=0.8$ (\%). Thus, these data are consistent with the possibility that the adult blood sample analyzed here is from the mother of the infant corpse.

Five genomic markers were also analyzed in DNA from the putative mother and infant corpse (vWA, TH01, CSF1PO, MCT118, and GPT) (data not shown). One genotype was detected in vWA and TH01 in the infant DNA sample, but two or three genotypes were observed in the other three markers. Thus, it appeared to be difficult to use genomic DNA to assess the relationship between the adult woman and the infant corpse.

In summary, this case study demonstrates that the method presented in this paper can be used to type mtDNA for forensic purposes. Because the input DNA may have been degraded in the case study described above, two rounds of PCR were performed with the same primers, using the same amount of DNA as in assays with undegraded DNA. With the second round of PCR, the risk of amplifying contaminating DNA increases; however, no DNA products were detected in a control sample lacking input DNA, suggesting that amplification of DNA contaminants did not occur. Amplification of short fragments also did not occur, despite the potentially degraded condition of the input DNA. Because this method was used successfully to type mtDNA from highly putrefied tissues, it is likely to be useful for forensic identification.

## References

1. Budowle B, Moretti TR, Niezgoda SJ, Brown BL. CODIS and PCR-based short tandem repeat loci: law enforcement tools. Genetic Identity Conference Proceedings of the Second European Symposium on Human Identification; 1998. Madison, WI: Promega Corporation, 1998;73-88; http:// www.promega.com/geneticidproc/eusymp2proc/default.htm
2. Gill P. Role of short tandem repeat DNA in forensic casework in the UKpast, present, and future perspectives. Biotechniques 2002;32:366-85.
3. Butler JM, Levin BC. Forensic applications of mitochondrial DNA. Trends Biotechnol 1998;16:158-62.
4. Holland MM, Parsons TJ. Mitochondrial DNA sequence analysis - varidation and use for forensic casework. Forensic Sci Rev 1999;11:21-50.
5. Budowle B, Allard MW, Wilson MR, Chakraborty R. Forensics and mitochondrial DNA: applications, debates, and foundations. Annu Rev Genomics Hum Genet 2003;4:119-41.
6. Greenberg BD, Newbold JE, Sugino A. Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. Gene 1983;21:33-49.
7. Lutz S, Weisser H-J, Heizmann J, Pollak S. A third hypervariable region in the human mitochondrial D-loop. Hum Genet 1997;101:384.
8. Piercy R, Sullivan KM, Benson N, Gill P. The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. Int J Legal Med 1993;106:85-90.
9. Sekiguchi K, Imaizumi K, Fujii K, Senju H, Mizuno N, Sakai I, et al. Sequence polymorphisms of the control region of human mitochondrial DNA in Japanese population (in Japanese). Jpn J Sci Technol Iden 1997;2:33-40.
10. Parson W, Parsons TJ, Scheithauer R, Holland MM. Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case. Int J Legal Med 1998;111:124-32.
11. Seo Y, Stradmann-Bellinghausen B, Rittner C, Takahama K, Schneider PM. Sequence polymorphism of mitochondrial DNA control region in Japanese. Forensic Sci Int 1998;97:155-64.
12. Yao Y-G, Kong Q-P, Bandelt H-J, Kivisild T, Zhang Y-P. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am J Hum Genet 2002;70:635-51.
13. Imaizumi K, Parsons TJ, Yoshino M, Holland MM. A new database of mitochondrial DNA hypervariable regions I and II sequences from 162 Japanese individuals. Int J Legal Med 2002;116:68-73.
14. Maruyama S, Minaguchi K, Saitou N. Sequense polymorphisms of the mitochondrial DNA control region and phylogenetic analysis of mtDNA lineages in the Japanese population. Int J Legal Med 2003;117:218-25.
15. Allard MW, Wilson MR, Monson KL, Budowle B. Control region sequences for East Asian individuals in the scientific working group on DNA analysis methods forensic mtDNA data set. Legal Med 2004;6:1124.
16. Lutz S, Wittig H, Weisser H-J, Heizmann J, Junge A, Dimo-Simonin N, et al. Is it possible to differentiate mtDNA by means of HVIII in samples that cannot be distinguished by sequencing the HVI and HVII regions? Forensic Sci Int 2000;113:97-101.
17. Lutz S, Weisser H-J, Heizmann J, Pollak S. mtDNA as a tool for identification of human remains. Identification using mtDNA. Int J Legal Med 1996;109:205-9.
18. Lutz S, Weisser H-J, Heizmann J, Pollak S. Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany [published errata appear in Int J Legal Med 1998;111:286 and 1999;112:145-50]. Int J Legal Med 1998;111:67-77.
19. Bini C, Ceccardi S, Luiselli D, Ferri G, Pelotti S, Colalongo C, et al. Different informativeness of the three hypervariable mitochondrial DNA regions in the population of Bologna (Italy). Forensic Sci Int 2003;135:4852.
20. Vanecek T, Vorel F, Sip M. Mitochondrial DNA D-loop hypervariable regions: Czech population data. Int J Legal Med 2004;118:14-8.
21. Zhang YJ, Xu QS, Zheng ZJ, Lin HY, Lee JB. Haplotype diversity in mitochondrial DNA hypervariable region I, II and III in northeast China Han. Forensic Sci Int 2005;149:267-9.
22. Penta JS, Johnson FM, Wachsman JT, Copeland WC. Mitochondrial DNA in human malignancy. Mutat Res 2001;488:119-33.
23. Chomyn A, Attardi G. MtDNA mutations in aging and apoptosis. Biochem Biophys Res Commun 2003;304:519-29.
24. Barnes WM. PCR amplification of up to $35-\mathrm{kb}$ DNA with high fidelity and high yield from $\lambda$ bacteriophage templates. Proc Natl Acad Sci USA 1994;91:2216-20.
25. Sambrook J, Russell DW. Molecular cloning, a laboratory manual. 3rd ed. New York: Cold Spring Harbor Laboratory Press, 2001.
26. Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. Nature 1981;290:457-65.
27. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 1999;23:147.
28. Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA. Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. Am J Hum Genet 1991;48:370-82.
29. Nei M, Tajima F. DNA polymorphism detectable by restriction endonucleases. Genetics 1981;97:145-63.
30. Kivisild T, Tolk H-V, Parik J, Wang Y, Papiha SS, Bandelt H-J, et al. The emerging limbs and twigs of the East Asian mtDNA tree [published erratum appears in Mol Biol Evol 2003;20:162]. Mol Biol Evol 2002;19:1737-51.
31. Kong Q-P, Yao Y-G, Sun C, Bandelt H-J, Zhu C-L, Zhang Y-P. Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences [published erratum appears in Am J Hum Genet 2004;75:157]. Am J Hum Genet 2003;73:671-6.
32. Mourier T, Hansen AJ, Willerslev E, Arctander P. The Human Genome Project reveals a continuous transfer of large mitochondrial fragments to the nucleus. Mol Biol Evol 2001;18:1833-7.
33. Bendall KE, Sykes BC. Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. Am J Hum Genet 1995;57:248-56.
34. Stewart JEB, Fisher CL, Aagaard PJ, Wilson MR, Isenberg AR, Polanskey D, et al. Length variation in HV2 of the human mitochondrial DNA control region. J Forensic Sci 2001;46:862-70.
35. Bär W, Brinkmann B, Budowle B, Carracedo A, Gill P, Holland M, et al. DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. Int J Legal Med 2000;113:1936.
36. Horai $S$, Hayasaka K. Intraspecific nucleotide sequence differences in the major noncording region of human mitochondrial DNA. Am J Hum Genet 1990;46:828-42.
37. Lee HY, Chung U, Park MJ, Yoo JE, Han GR, Shin KJ. Differential distribution of human mitochondrial DNA in somatic tissues and hairs. Ann Hum Genet 2005;70:59-65.
38. Bini C, Pappalardo G. mtDNA HVI length heteroplasmic profile in different tissues of maternally related members. Forensic Sci Int 2005;152:35-8.
39. Ogata M, Mattern R, Schneider PM, Schacker U, Kaufmann T, Rittner C. Quantitative and qualitative analysis of DNA extracted from postmortem muscle tissues. Z Rechtsmed 1990;103:397-406.
40. Atmadja DS, Tatsuno Y, Ueno Y, Nishimura A. The effect of extraction methods, the kind of organ samples and the examination delay on the DNA yields and typing. Kobe J Med Sci 1995;41:197-211.

Additional information and reprint requests:
Tadashi Mabuchi, Ph.D.
Department of Biochemistry
Interdisciplinary Graduate School of Medicine and Engineering
University of Yamanashi
1110 Shimokato, Chuo, Yamanashi 409-3898
Japan
E-mail: tmabuchi@yamanashi.ac.jp


[^0]:    ${ }^{1}$ Department of Biochemistry, Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, 1110 Shimikato, Chuo, Yamanashi 409-3898, Japan.
    ${ }^{2}$ Department of Legal Medicine, Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, 1110 Shimokato, Chuo, Yamanashi 409-3898, Japan.

    Received 11 Dec. 2005; and in revised form 30 April and 26 Aug. 2006; accepted 16 Sept. 2006; published 30 Jan. 2007.

[^1]:    Defining characters for haplogrouping are indicated by bold and italic, and those in the cording region are indicated by parentheses.
    $\mathrm{Cn}, n+1, n+2, n+3$ in C -stretches indicate heteroplasmies composed with $\mathrm{Cn}, n+1, n+2$ and $n+3$.
    [Correction added after publication 30 January 2007: in the preceding table missing data was added.]

[^2]:    Heteroplasmic sequences are indicated by a shadowed region: components in each heteroplasmic variant are indicated by a vertical line.

