

Mitochondrial Sequence Analysis for Forensic Identification Using Pyrosequencing Technology

BioTechniques 32:124-133 (January 2002)

**H. Andréasson, A. Asp¹,
A. Alderborn¹, U. Gyllensten,
and M. Allen**

Uppsala University, and
¹Pyrosequencing AB,
Uppsala, Sweden

ABSTRACT

Over recent years, requests for mtDNA analysis in the field of forensic medicine have notably increased, and the results of such analyses have proved to be very useful in forensic cases where nuclear DNA analysis cannot be performed. Traditionally, mtDNA has been analyzed by DNA sequencing of the two hypervariable regions, HVI and HVII, in the D-loop. DNA sequence analysis using the conventional Sanger sequencing is very robust but time consuming and labor intensive. By contrast, mtDNA analysis based on the pyrosequencing technology provides fast and accurate results from the human mtDNA present in many types of evidence materials in forensic casework. The assay has been developed to determine polymorphic sites in the mitochondrial D-loop as well as the coding region to further increase the discrimination power of mtDNA analysis. The pyrosequencing technology for analysis of mtDNA polymorphisms has been tested with regard to sensitivity, reproducibility, and success rate when applied to control samples and actual casework materials. The results show that the method is very accurate and sensitive; the results are easily interpreted and provide a high success rate on casework samples. The panel of pyrosequencing reactions for the mtDNA polymorphisms were chosen to result in an optimal discrimination power in relation to the number of bases determined.

INTRODUCTION

In recent years, rapid developments in the area of molecular genetics have made it possible to perform DNA analysis on materials such as blood-stains, semen stains, and other biological materials as evidence in forensic casework (5,21). The sensitivity of the PCR method has even made it possible to analyze very old and degraded DNA material (10,13,19). Moreover, the ability to analyze DNA by PCR has made it possible to type a wider range of biological material found at the scene of a crime. Saliva stains, teeth, skeletal remains, shed hairs (4,11,12), skin cells (15), and fingerprints (30) on various items are frequently used as evidence materials today. In fact, a majority of the evidence materials found at a crime scene contains sufficient amounts of DNA for genetic analysis.

Most of the laboratories worldwide that perform routine casework analysis have focused on fragment size analysis of multiplex short tandem repeat (STR) loci. The STR analysis is highly discriminating and well established as a routine forensic diagnostic tool. The large National DNA databases with STR profiles from crime scene materials and convicted felons provide a major advantage in nuclear DNA analysis. However, there are several evidence samples where the nuclear DNA is too scarce to perform a multi-locus STR analysis. In these cases, mtDNA can often be used as an alternative genetic source because of the high number of copies per cell. Sequencing of mtDNA has proven very useful for criminal investigations where the evidence materials have been in very small amounts or severely degraded, such as in hair shafts and bones (3). In routine mtDNA analysis, HVI and HVII, the two most hypervariable regions of the control re-

gion (the D-loop), are sequenced. The method has proven to be extremely sensitive, with the potential to amplify and sequence 30 fg DNA or approximately 10 mtDNA molecules (3). This analysis has been tested and further evaluated on a variety of materials containing minute amounts of DNA in our laboratory. Items such as shed hair, jewelry, spectacles, shoes, clothes, wigs, fake moustaches, baseball caps, and cellular phones are often found at crime scenes, and the DNA analysis of these materials has proven very valuable in forensic investigations. A major drawback of using conventional sequencing (28) of the D-loop is the time consuming procedure. For example, a case with 4–6 samples may take several weeks to complete because several tests have to be performed sequentially from each sample to safeguard results.

The extended variation in mtDNA among individuals has been studied for forensic analysis using mainly Sanger sequencing. Sanger sequencing analysis is an elegant but labor-intensive method involving multiple steps and an electrophoretic slab-gel separation of DNA fragments. Methods that are less labor intensive have been developed, such as sequence-specific oligonucleotide (SSO) analysis hybridization (29), minisequencing (16), and immobilized SSO probes (22). These methods are easy to perform in a few hours, but their discrimination power is substantially lower in comparison with a sequence analysis. These systems are designed for analysis of 12 to 23 of the most polymorphic sites in the mtDNA D-loop rather than the full sequence resulting in a lower probability of a random match. Consequently, these systems are used mostly for pre-screening to exclude mtDNA identity between samples. Whenever the pre-screening analysis indicates an mtDNA match between

two samples, a complete sequence analysis is used to further confirm the identity between the samples. With the development of the rapid high-throughput technologies for genetic analysis such as DNA microarrays and pyrosequencing, alternative methods for forensic identification are emerging (20,27).

PyrosequencingTM was first developed for expressed sequence tag sequencing and sequencing of short stretches of DNA but has mainly been used for detection of known single-nucleotide polymorphisms (SNPs) (17,24, 27). In comparison to many other techniques for SNP analysis, pyrosequencing is very quick, robust, easy to use, and has a high-throughput capability (2,24). The actual SNP genotyping reaction of 96 samples is performed in less than 10 min (1). Pyrosequencing is a non-electrophoretic single-tube DNA sequencing method that uses an enzyme cascade system to monitor DNA synthesis. The technology is based on the detection of released pyrophosphate (PP_i) during DNA synthesis (24–26). Nucleotides are added one at a time to the reaction, and if the nucleotide is incorporated, pyrophosphate is released and converted to ATP by ATP-sulfurylase. Luciferase uses the ATP to generate detectable light. The amount of light generated is proportional to the number of nucleotides incorporated. If the added nucleotide does not form a base pair to the DNA template, then the polymerase will not incorporate it and no light will be produced. Instead, the nucleotide will be degraded rapidly by apyrase. The addition of successive rounds of nucleotides and the detection of emitted light are performed in an automated instrument. This new method for the detection of known SNPs or short DNA sequences is gaining widespread use in many areas.

The turnaround time for the forensic analysis of mtDNA can be reduced substantially using pyrosequencing to determine the mtDNA sequence. Further developments in pyrosequencing technology [sequence analysis (SQA)] have made it possible to easily determine short stretches (50–100 nucleotides) of a DNA sequence. Sequence analysis has proven to be successful on fragments up to 109 bp of the p53 gene in one single pyrosequencing reaction (9).

Even though the analysis still has to be repeated several times, the total analysis time can be reduced from several weeks to a few days. This will allow the use of an mtDNA analysis during the first important days of a criminal investigation.

Although the discrimination power is far from what can be obtained when analyzing multiplex nuclear markers, D-loop analysis provides a very useful tool for the DNA analysis of limited DNA samples. A way to increase the discrimination power of forensic mtDNA analysis is to analyze polymorphisms in the coding region in addition to the D-loop region. There is a lack of completely analyzed mitochondrial genomes submitted to public databases that can be used to select informative SNPs. Therefore, a database of completely sequenced mitochondrial genomes from 26 individuals of Swedish origin has been established.

This pyrosequencing assay, developed for forensic mtDNA determination, is based on the analysis of 13 PCR fragments, covering the most informative sites in the entire mitochondrial genome. The system relies on the analysis of a total of 19 pyrosequencing reactions, 4 HVI, 4 HVII, and 11 coding region reactions. This will allow the analysis of the most discriminating regions of the entire mtDNA genome for forensic identification using a PSQTM 96 System (Pyrosequencing AB, Uppsala, Sweden). A forensic typing assay enabling the analysis of the D-loop and parts of the mitochondrial coding region is suitable for the routine analysis of a wide range of forensic materials with limited and degraded DNA. Moreover, the turnaround time is shorter and the discrimination power is higher in comparison to conventional D-loop sequencing alone.

MATERIALS AND METHODS

DNA Preparations

Human DNA was extracted from peripheral blood lymphocytes of 190 Swedish blood donors to serve as control samples. The forensic evidence materials were purified by one of three methods. The Wizard[®] Genomic DNA Extraction Kit (Promega, Madison, WI, USA) was used to extract DNA from

most of the evidence materials collected by cotton swabs and blood. Chelex[®] 100 (Bio-Rad Laboratories, Hercules, CA, USA) was used to extract DNA from bloodstains. This method was developed for extracting DNA from forensic samples (32). Hair samples were extracted using an extraction procedure based on proteinase K and DTT (31).

Primer Design and PCR

The software Primer ExpressTM (Applied Biosystems, Foster City, CA, USA) was used for PCR and sequence primer design. An optimal target selection was made to cover most of the informative polymorphism in the D-loop. On the basis of the determined heterogeneity in the Swedish and other populations (14), a number of informative regions outside the D-loop were chosen for analysis in addition to the D-loop. These regions are highly polymorphic, with SNP frequencies in a range between 0.15 and 0.50, serving as suitable mtDNA markers for forensic identification.

HVI and HVII were PCR amplified in two separate reactions. Each of the two templates was analyzed using four sequencing primers per fragment, including the forward PCR primer. A large number of the polymorphisms in the D-loop will be determined based on these eight short overlapping sequences. The coding region fragments were amplified in 11 separate reactions, followed by analysis in 11 pyrosequencing reactions using the forward PCR primer as sequencing primer. Each of the regions outside the D-loop contains from two to eight polymorphic sites within the range of nucleotides (50 nucleotides) that can be read by the SQA reaction. Control-sample PCRs contain 1.5 μ L DNA, 200 nM of each primer (Table 1), 200 μ M of each dNTP, 1.5 mM MgCl₂, 2 U AmpliTaq Gold[®] DNA polymerase, and 1 \times GeneAmp[®] PCR Buffer II (Applied Biosystems) in a total volume of 70 μ L. PCR amplifications for the forensic materials contain 10 μ L DNA, 200 nM of each primer, 200 μ M of each dNTP, 2.4 mM MgCl₂, 10 U AmpliTaq Gold DNA polymerase, 1.2 \times GeneAmp PCR Buffer II, 0.16 mg/mL BSA, and 10% glycerol in a total volume of 100 μ L. The amplifications were performed in a

Research Report

Table 1. PCR and Sequencing Primers

| Primer Name ^a | PCR Primer | Sequence Primer | T _m (°C) | Sequence (5'→3') | Dispensation Order | SSB Required |
|--------------------------|------------|-----------------|---------------------|----------------------------|--------------------|--------------|
| II 45 | X | X | 55.6 | ATGCATTTGGTATTTTCGTCTG | TCGA | |
| II 111 | | X | 48.2 | ACCCTATGTCGCAGTATCT | TCGA | |
| II 162 | | X | 54.4 | CGCACCTACGTTCAATATTACA | CTGA | |
| II 216 | | X | 48.3 | TTAATGCTTGTAGGACATAATAA | CTGA | |
| II 287 B | X | | 55.6 | TTGTTATGATGTCTGTGTGGAAAG | | |
| C 431 | X | X | 53.6 | CACCCCCCAACTAACACA | ACGT | |
| C 637 B | X | | 53.1 | GTGATGTGAGCCCGTCTAA | | |
| C 2988 | X | X | 51.8 | CGATGTTGGATCAGGACA | ACGT | |
| C 3216 B | X | | 52.4 | GGGTGGGTGTGGGTATAA | | |
| C 3403 | X | X | 56.8 | CTACGCAAAGGCCCAA | CAGT | X |
| C 3641 B | X | | 56 | GCTAGGCTAGAGGTGGCTAGAA | | |
| C 4156 | X | X | 52.5 | CAACTCATAACCTCCTATGAAA | ACGT | |
| C 4367 B | X | | 52.6 | TTGGATTCTCAGGGATGG | | |
| C 4882 | X | X | 49 | CCATCTCAATCATATACCAA | ATGC | |
| C 5138 B | X | | 50 | GGAGTTTAAGTTGAGTAGTAGGAA | | |
| C 8665 | X | X | 52 | CAATGACTAATCAAACCTCA | ATGC | |
| C 8803 B | X | | 51.1 | TAAATGAGTGAGGCAGGAGT | | |
| C 12346 | X | X | 47.9 | CACACTACTATAACCACCCTAA | ACGT | X |
| C 12541 B | X | | 49.1 | CTCAGTGTGAGTTGAGATAA | | |
| C 12673 | X | X | 45.7 | AACATTAATCAGTTCTTCAA | ACGT | X |
| C 12861 B | X | | 46.7 | GTTGTATAGGATTGCTTGAA | | |
| C 14747 | X | X | 53.4 | ATGACCCCAATACGCAA | ACGT | X |
| C 14949 B | X | | 54.1 | TGGGCGATTGATGAAAAG | | |
| C 15883 | X | X | 50.2 | GGCCTGTCCTTGTAGTATAAA | ACGT | |
| C 16083 B | X | | 52 | GGTTGTTGATGGGTGAGTC | | |
| C 16496 | X | X | 48.5 | GACATCTGGTTCCTACTTCA | ACGT | X |
| C149 B | X | | 47.7 | ATGAGGCAGGAATCAA | | |
| I 16105 | X | X | 55.4 | TGCCAGCCACCATGAATA | CTGA | X |
| I 16168 | | X | 50.3 | CCAATCCACATCAAAACC | CTGA | X |
| I 16203 | | X | 45.7 | AGCAAGTACAGCAATCAA | CTGA | |
| I 16266 | | X | 47.3 | CCCCTAGGATACCAACA | CTGA | |
| I 16348 B | X | | 55.6 | GACTGTAATGTGCTATGTACGGTAAA | | |

^aI, HVI; II, HVII; C, coding region; B, biotin-labeled reverse primer.

GeneAmp PCR System 9600 (Applied Biosystems). The samples were kept for 10 min at 95°C, followed by 40 cycles of 30 s at 95°C, 45 s at 60°C (53°C for the coding region primers), 60 s at 72°C, with a final extension step for 7 min at 72°C.

Template Preparation and Pyrosequencing Reaction

Streptavidin-coated beads were used as a solid-phase support to obtain single-stranded biotinylated PCR products, as described by Pyrosequencing AB.

When necessary, 440 ng Single-Stranded DNA Binding Protein[®] (SSB) (Amersham Pharmacia Biotech AB, Uppsala, Sweden) were added to the primed DNA template before pyrosequencing (23). The sequencing was performed at room temperature, and 15 µL

Research Report

of the PCR were used with 400 nM sequence primer. Enzyme and substrate (prototype of PSQ 96 SQA Reagent Kit) were added to all samples in a PSQ 96 System with a prototype of PSQ 96 SQA Software (sample entry, instrument control, and evaluation). The procedure was carried out by stepwise elongation of the primer strand during the sequential dispensation of different dNTPs (AαS, C, G, and T), followed by the degradation of the nucleotides. Optimal cyclic dispensation orders were chosen for each fragment. Some sequences required manual editing because of incorrect base calling by the SQA software when scoring homopolymeric peaks before alignment to the Anderson et al. (6) reference sequence. For comparison, control samples were sequenced using an ABI PRISM[®] 377 instrument and BigDye[™] Terminator chemistry (Applied Biosystems).

RESULTS

The analysis of the mitochondrial D-loop was performed using two separate PCR fragments. The HVI and HVII fragments were analyzed in eight separate pyrosequencing reactions. To evaluate the pyrosequencing technology in forensic analysis, a number of control samples were analyzed. In total, 190 samples were analyzed for HVII, 120 samples were analyzed for HVI, and, finally, 50 forensic evidence materials were analyzed for HVII. The results have been compared with D-loop sequencing data from these individuals.

The pyrosequencing of the HVII reactions yielded a read length of more than 30 nucleotides in the sequence analysis of the control samples without manual editing and more than 50 nucleotides in general after editing. Among control samples and evidence materials, the HVII sequencing reactions could be read satisfactorily without any further optimization. The HVI reactions required further optimization to reduce the background and extend the read length. The addition of the SSB to the HVI reactions extended the read length up to 27 nucleotides before editing and 50 nucleotides in general after editing (Table 2). The longest read length, 71 nucleotides, was obtained

when analyzing the HVII reaction II 162. The shortest read lengths, 45 nucleotides after editing, were seen in the HVII 216 and HVI 16105 reactions (Table 2). Some of the reactions show a large difference in the number of nucleotides that can be read before and after editing. The difference is due to problems with reading the number of incorporated nucleotides in the homopolymeric regions. This creates a frame shift in comparison with the reference sequence but can easily be overcome by manually editing the sequence. The inherent problem to sequence through homopolymeric regions has been shown to be reduced by the addition of SSB in the pyrosequencing reaction. The use of SSB results in longer read lengths, increased signal intensity, and more accurate base-calls in homopolymeric regions (23).

The results of the eight HVI and HVII pyrosequencing analyses were aligned and compared with the reference D-loop sequence to evaluate the coverage in the specific region. Approximately 69% (153/222 nucleotides) of the HVII region was covered before editing and 95% (211/222 nucleotides) after editing of the sequences. Similarly, 42% (96/226 nucleotides) of the HVI region was covered before editing and 82% (185/226 nucleotides) after editing. However, the lack of complete coverage will only result in the exclusion of three known polymorphic sites for HVII within the 222 nucleotides in this analysis. These polymorphisms are not very common among our Swedish control samples, showing a frequency of 1/355 individuals for two of the polymorphisms and 8/355 for the third polymorphism. After editing the sequence, 84% of the HVI region was covered. This results in a loss of nine polymorphic sites in the analysis. These polymorphisms are somewhat more common than the uncovered sites in HVII among our control samples (ranging from 1/318 to 31/318). All HVII and HVI sequences, from both control samples and evidence material, could be confirmed to 100% (120/120 individuals) when compared with data from previous termination sequencing of the D-loop.

Pyrosequencing was further used to sequence 11 mtDNA coding regions in 36 previously sequenced control sam-

ples. The 11 regions were chosen to cover the most informative sites throughout the entire mitochondrial genome using the available complete mtDNA sequences. The results obtained using pyrosequencing were 100% identical to those obtained by the Sanger sequencing method. Also here, some of the reactions required the use of SSB to obtain longer read lengths (Table 3). The analysis of the coding region yielded 12 to 39 nucleotides before editing and generally 50 to 60 nucleotides after editing. The addition of SSB improved the read length for some fragments and reduced the background signal for other fragments. The longest read length was obtained in the analysis of fragment C 12346, where 67 nucleotides could be read. The shortest read length was seen for the fragment C 431, where only 31 nucleotides could be read. In addition, the fragment C 8665 revealed only 35 nucleotides in the analysis (Table 3). However, this fragment was analyzed, limiting the number of dispensations to 60, in contrast to the 96 that were normally used. Allowing 96 dispensations for this fragment will probably extend the read length up to 20 additional bases. In the pyrosequencing analysis of all coding region fragments, there were a total of 20 different SNPs detected among 36 individuals. Each coding region fragment was found to contain between 0 and 4 different SNPs (Table 3). Among all SNPs five (in four different fragments) were detected in frequencies between 19% and 50%.

To evaluate the system on limited amounts of DNA, 50 actual casework samples previously studied by termination sequencing were analyzed by pyrosequencing. Among the evidence materials analyzed using pyrosequencing were robber hoods, wigs, moustaches, shoes, mobile phones, watches, knives, guns, and fingerprints. In addition, saliva from very old letters (1953) and paraffin-embedded tissue were analyzed. Results using pyrosequencing on the forensic samples were completely comparable with the results obtained using termination sequencing. Pyrosequencing of HVII yielded, on average, more than 30 nucleotides in the analysis of the old evidence materials without manual editing and more than 50 nucleotides after editing.

Research Report

Furthermore, pyrosequencing analysis was used in a forensic case (Figures 1 and 2, and Table 4). The case involved the brutal murder of an elderly woman. During the investigation, two shed hairs were found on the victim. Analysis of the two hairs showed that neither of the hairs could have come from the suspect.

DISCUSSION

There is a strong need to establish efficient technology platforms in the area of forensic analysis to meet the requirement for the rapid and reliable identification of the origin of diverse biological materials. The described analysis of mtDNA for identification purposes based on pyrosequencing technology meets these requirements. Pyrosequencing has proven to be successful for the sequence analysis of the two PCR fragments covering the two highly informative HVI and HVII regions of the D-loop and several regions

Table 2. Number of Read Nucleotides in Sequence Analysis of the D-Loop Region in the Control Samples

| Fragment/Primer | Maximum Read Length (Nucleotides) | Read Nucleotides Before Editing | Read Nucleotides After Editing |
|-----------------|-----------------------------------|---------------------------------|--------------------------------|
| II 45 | 57 | 45 | 57 |
| II 111 | 65 | 32 | 63 |
| II 162 | 71 | 43 | 69 |
| II 216 | 47 | 33 | 45 |
| I 16105 | 50 | 25 | 45 |
| I 16168 | 50 | 27 | 50 |
| I 16203 | 54 | 22 | 53 |
| I 16266 | 53 | 22 | 52 |

An average estimation based on the analysis of 96 individuals.

outside the D-loop. A large number of blood donor samples and forensic evidence samples have successfully been analyzed. Most of the pyrosequencing reactions read routinely between 40 and 50 bp with high accuracy.

Combining the results of the eight HVI and HVII pyrosequencing reactions altogether cover approximately 56% (249/448 nucleotides) before editing and 88% (396/448 nucleotides) after manual editing of the PCR frag-

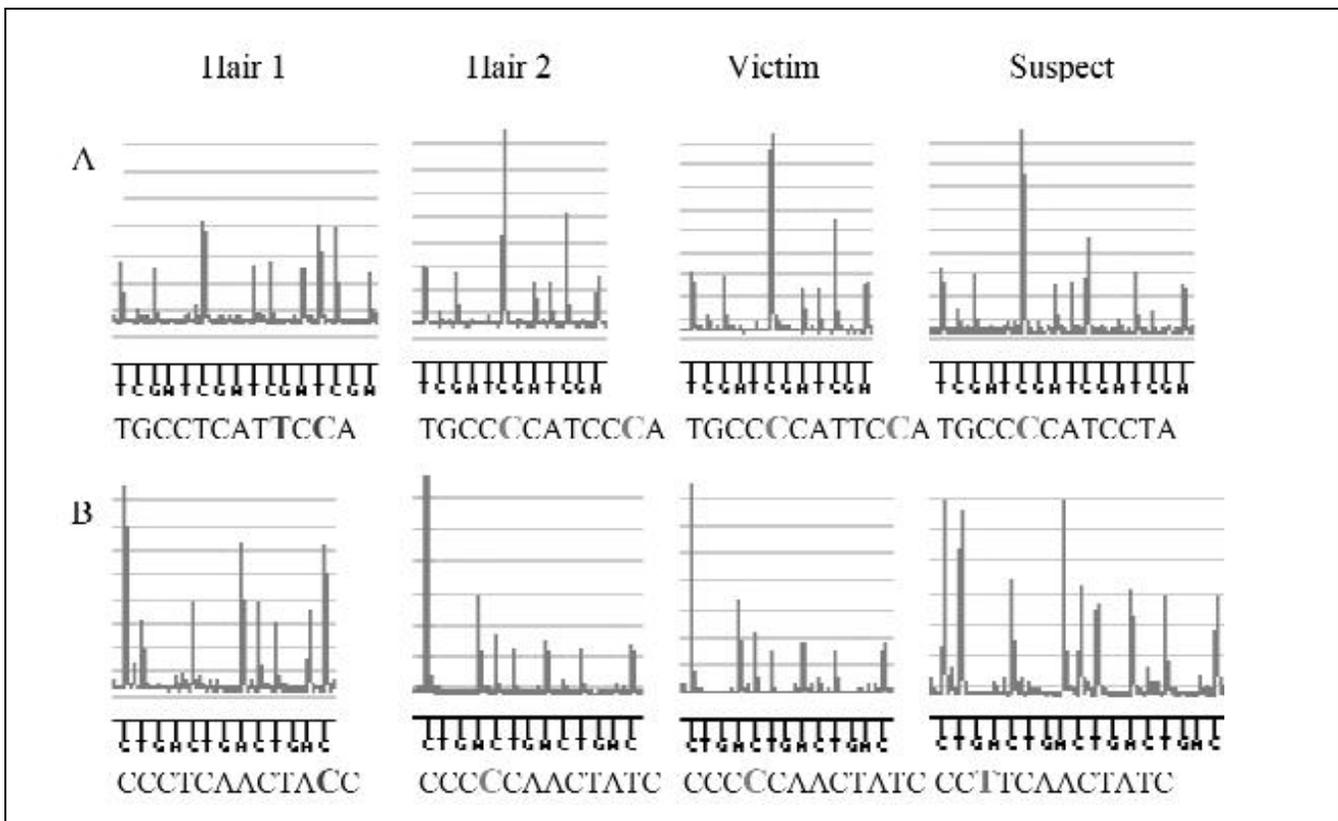


Figure 1. Results from pyrosequencing analysis of evidence materials in casework. (A) Polymorphisms revealed at nucleotides 146, 150, and 152. (B) Polymorphisms revealed at nucleotides 16223, 16224, and 16231.

Table 3. Number of Read Nucleotides in Sequence Analysis of the Coding Regions in the Control Samples

| Fragment/ Primer | SNPs per Fragment ^a Sw ^b /Ingman et al. ^c | Before Editing | After Editing | SNPs Detected by Pyrosequencing |
|---------------------|---|-------------------|------------------|------------------------------------|
| C 431 | 4/4 | 12 | 31 | 2 |
| C 2988 | 1/1 | 29 | 64 | 1 |
| C 3403 | 3/1 | 24 | 59 | - |
| C 4156 | 2/1 | 17 | 53 | 2 |
| C 4882 | 2/- | 39 | 56 | 1 |
| C 8665 | 3/1 | 10 | 35 ^d | 2 |
| C 12346 | 2/1 | 27 | 67 | 1 |
| C 12673 | 2/1 | 38 | 66 | 2 |
| C 14747 | 3/3 | 21 | 60 | 3 |
| C 15883 | 5/4 | 31 | 65 | 4 |
| C 16496 | 2/3 | 27 | 62 | 2 |

The read length is an average estimate in 36 samples.

^aNumber of SNPs found within 50 bp from forward primer.

^bTwenty whole mitochondrial DNA-sequenced control samples from individuals of Swedish origin.

^cFifty-three whole mitochondrial DNA-sequenced control samples from individuals in other populations (14).

^dAverage value is based on 60 dispensations instead of 96.

ments. Complete Sanger sequencing of the D-loop is usually done on a 359-bp HVII fragment and a 403-bp HVI fragment using the primer pairs L048/H408 and L15997/R16401 (33). The pyrosequencing assay will, as designed here, cover 62% (222/359 nucleotides) of the HVII fragment determined by sequencing and 56% (226/403 nucleotides) of the HVI fragment determined by sequencing. In total, this gives 59% coverage (448/762) of the D-loop determined by conventional sequencing. Although the coverage is less for the pyrosequencing assay in comparison to conventional Sanger sequencing, the majority of the polymorphisms found in our Swedish database are located within the area studied by pyrosequencing (nucleotides 66–287 and 16123–16348). The analysis of the HVI region resulted in a few positions within this area not covered by the pyrosequencing reactions. The most common of them (31/318) might be covered by one of the sequencing primers if the

Table 4. Results from the Analysis of Forensic Evidence Materials in a Forensic Investigation Using Pyrosequencing

| Primer Name | Nucleotide Position ^a | Evidence | | Reference | |
|----------------|-------------------------------------|----------------|----------------|-----------|---------|
| | | Hair 1 | Hair 2 | Victim | Suspect |
| II 45 | 73 (A) | G | G | G | G |
| II 111 | 146 (T) | T | C | C | C |
| | 150 (C) | T | C | C | C |
| | 152 (T) | C | C | C | T |
| II 162 | 195 (T) | C | T | T | T |
| II 216 | 263 (A) | G | G | G | G |
| C 12346 | 12372 (G) | A | A | A | G |
| C 12672 | 12705 (C) | C | C | C | T |
| C 16496 | 16519 (T) | T | Not determined | C | C |
| I 16105 | 16126 (T) | C | T | T | T |
| | 16145 (G) | A | G | G | G |
| I 16168 | 16189 (T) | T | T | T | C |
| I 16203 | 16223 (C) | C | C | C | T |
| | 16224 (T) | T | C | C | T |
| | 16231 (T) | C | T | T | T |
| | 16261 (C) | Not determined | C | C | T |
| I 16266 | 16311 (T) | T | C | C | C |
| | 16319 (G) | G | G | G | A |

An identical mtDNA type was seen between Hair 2 and the victim, while Hair 1 and the suspect have unique mtDNA types.

^aAnderson et al. (6) reference sequence within brackets.

Research Report

system is optimized further. Also, a re-design of some of the sequencing primers might solve this problem. Further, the coding region fragments were analyzed with high accuracy in the analysis of control samples. All fragments were read between 31 and 67 nucleotides after editing, and the majority of the fragments could routinely be read over 50 nucleotides. Overall, 20 different SNPs were detected within the 11 coding region fragments, of which some were very frequent.

Different cyclic dispensation orders of dNTPs were found to be suitable for different regions (Table 1). The read lengths were extended and the background was reduced by choosing optimal dispensation orders. A manually programmed dispensation order adapted to the analyzed sequence can increase the read length even further. This is done by entering the order of the nucleotides expected from the reference sequence with additional nucleotides in positions where polymorphisms are likely to occur (9).

An important issue in forensic analy-

sis based on the analysis of mtDNA is heteroplasmy (8). Although heteroplasmy is relatively rare, misinterpretations can be made if appropriate precautions are not taken (7). A major advantage in the analysis of heteroplasmy and DNA mixtures is the ability to quantify specific amounts of the two separate DNA types in a sample. An assay based on pyrosequencing has been used to quantify virus copy number in the presence of a competitor (18). Moreover, pyrosequencing has been used to determine population frequencies for SNPs in large pools of samples (Application note; <http://www.pyrosequencing.com/pages/applications.html>). Although the technology has been used for quantification purposes previously, further investigations are necessary to evaluate pyrosequencing on heteroplasmic and mixed forensic samples.

In cases where the mtDNA type found in the evidence materials matches the mtDNA type of a suspect, it is necessary to estimate the frequency distribution of that particular mtDNA sequence in the relevant population. A

database of complete D-loop sequences from 350 Swedish individuals has been established. Furthermore, the frequency of the informative sites in the coding regions will be determined in the Swedish population. Here, the pyrosequencing method can be used for a rapid generation of reference databases. The estimation of the heterogeneity in the coding region among the 350 control individuals will further be used to optimize the assay. The eight most informative regions in the coding region will be chosen for analysis in routine forensic case-work. Thus, a complete analysis will include 16 pyrosequencing reactions, eight reactions for the D-loop and eight reactions for the coding region.

A major problem in forensic analysis is the limited amount of DNA available, and it is therefore preferable to limit the number of reactions to a minimum. Forensic material is also very often degraded and may contain inhibitors that will lower the efficiency of the reaction. In view of this, a multiplex PCR system is under development. The analysis will be performed on the products from two to three multiplex PCRs rather than 13 separate reactions. The pyrosequencing assay for mtDNA analysis is very flexible, and other polymorphic markers suitable for forensic analysis can easily be added in the future. Such additional markers can include mitochondrial or nuclear markers, either as single SNPs or sets of SNPs within a short stretch of DNA. To a large extent, this assay can be used instead of the conventional Sanger sequence analysis in forensic investigations, and the high sensitivity makes it well suited for the analysis of most of the evidence samples found at crime scenes. In addition, one or several pyrosequencing reactions can be used to pre-screen samples to reduce the number of samples that require a full sequence analysis to a minimum. Generally, mtDNA analysis is employed mainly in severe, violent forensic cases. Further developments of the mtDNA analysis may stimulate the use of this method in other categories of cases as well. The technology can, to a large extent, be automated and may in the future be developed into a fully automated system that minimizes the errors caused by sample mix-ups or pipetting variation.

In conclusion, the pyrosequencing

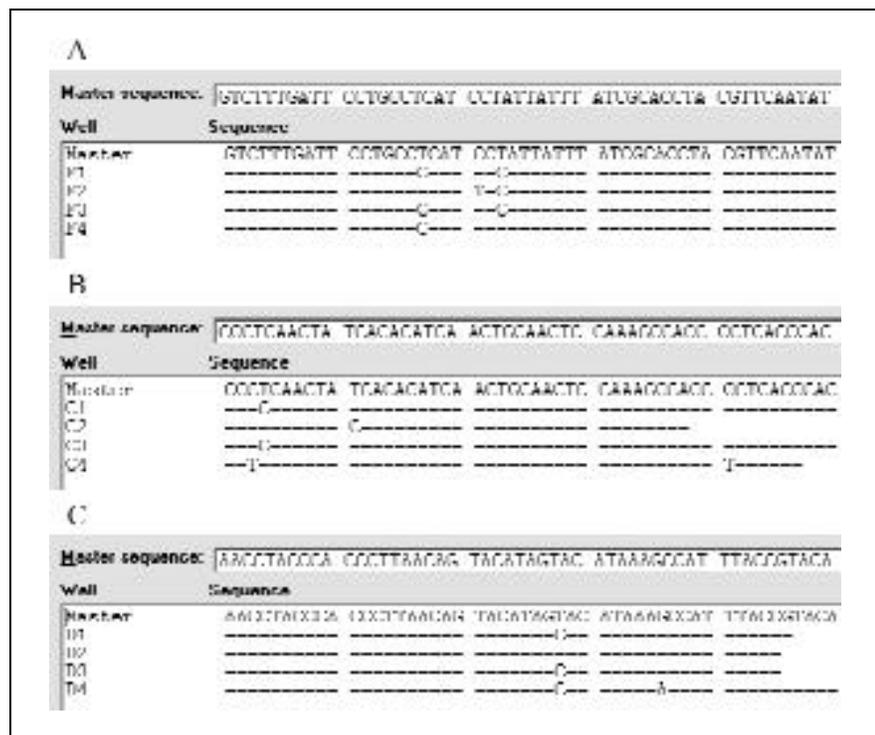


Figure 2. Alignment of pyrosequencing results for three of the reactions in a forensic case. Comparison is done using SQA Software to evaluate the pyrosequencing analysis. F1, C1, D1 = hair 1; F2, C2, D2 = hair 2; F3, C3, D3 = the victim; and F4, C4, D4 = the suspect. Master sequence = Anderson et al. (6) reference sequence. (A) Nucleotides 146, 150, and 152 are showing differences. (B) Nucleotides 16223, 16224, 16231, and 16261 are showing differences. (C) Nucleotides 16311 and 16319 are showing differences.

assay for sequence analysis of the most informative sites in the entire mitochondrial genome is well suited for forensic DNA analysis. The final system will be based on the analysis of a total of 16 pyrosequencing reactions to determine the polymorphisms in the D-loop and coding region allowing analysis of six samples in a 96-well system or 24 samples in a 384-well system simultaneously. Even if the analysis of the casework samples has to be performed at two independent occasions to exclude errors introduced by the operator or procedure, the analysis time can be reduced substantially from several weeks to a few days using pyrosequencing as compared to alternative methods. Thus, the mtDNA pyrosequencing analysis is much faster than the traditional sequencing of mtDNA performed today while still retaining a comparable resolution of the existing sequence polymorphism. In addition, the ability to analyze polymorphisms in the coding region may increase the discrimination power in casework analysis.

ACKNOWLEDGMENTS

This work was supported by grants from the Swedish National Board for Industrial and Technical Development (NUTEK), the Beijer Foundation, and the Swedish Research Council for Medical Sciences (MFR).

REFERENCES

- Ahmadian, A., B. Gharizadeh, A.C. Gustafsson, F. Sterky, P. Nyren, M. Uhlen, and J. Lundeberg. 2000. Single-nucleotide polymorphism analysis by pyrosequencing. *Anal. Biochem.* 280:103-110.
- Alderborn, A., A. Kristofferson, and U. Hammerling. 2000. Determination of single-nucleotide polymorphisms by real-time pyrophosphate DNA sequencing. *Genome Res.* 10:1249-1258.
- Allen, M., A.S. Engstrom, S. Meyers, O. Handt, T. Saldeen, A. von Haeseler, S. Paabo, and U. Gyllensten. 1998. Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: sensitivity and matching probabilities. *J. Forensic Sci.* 43:453-464.
- Allen, M., T. Saldeen, and U. Gyllensten. 1994. PCR-based DNA typing of saliva on stamps and envelopes. *BioTechniques* 17:546-552.
- Allen, M., T. Saldeen, U. Pettersson, and U. Gyllensten. 1993. Genetic typing of HLA class II genes in Swedish populations: application to forensic analysis. *J. Forensic Sci.* 38:554-570.
- Anderson, S., A.T. Bankier, B.G. Barrell, M.H. de Bruijn, A.R. Coulson, J. Drouin, I.C. Eperon, D.P. Nierlich et al. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290:457-465.
- Bar, W., B. Brinkmann, B. Budowle, A. Carracedo, P. Gill, M. Holland, P.J. Lincoln, W. Mayr et al. 2000. DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. *Int. J. Legal Med.* 113:193-196.
- Calloway, C.D., R.L. Reynolds, G.L. Herrin, Jr., and W.W. Anderson. 2000. The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. *Am. J. Hum. Genet.* 66:1384-1397.
- Garcia, C.A., A. Ahmadian, B. Gharizadeh, J. Lundeberg, M. Ronaghi, and P. Nyren. 2000. Mutation detection by pyrosequencing: sequencing of exons 5-8 of the p53 tumor suppressor gene. *Gene* 253:249-257.
- Gill, P., P.L. Ivanov, C. Kimpton, R. Piercy, N. Benson, G. Tully, I. Evett, E. Hagelberg et al. 1994. Identification of the remains of the Romanov family by DNA analysis. *Nat. Genet.* 6:130-135.
- Ginther, C., L. Issel-Tarver, and M.C. King. 1992. Identifying individuals by sequencing mitochondrial DNA from teeth. *Nat. Genet.* 2:135-138.
- Hagelberg, E., B. Sykes, and R. Hedges. 1989. Ancient bone DNA amplified. *Nature* 342:485.
- Handt, O., M. Richards, M. Trommsdorff, C. Kilger, J. Simanainen, O. Georgiev, K. Bauer, A. Stone et al. 1994. Molecular genetic analyses of the Tyrolean Ice Man. *Science* 264:1775-1778.
- Ingman, M., H. Kaessmann, S. Paabo, and U. Gyllensten. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708-713.
- Lorente, M., C. Entrala, J.A. Lorente, J.C. Alvarez, E. Villanueva, and B. Budowle. 1998. Dandruff as a potential source of DNA in forensic casework. *J. Forensic Sci.* 43:901-902.
- Morley, J.M., J.E. Bark, C.E. Evans, J.G. Perry, C.A. Hewitt, and G. Tully. 1999. Validation of mitochondrial DNA minisequencing for forensic casework. *Int. J. Legal Med.* 112:241-248.
- Nordstrom, T., B. Gharizadeh, N. Pourmand, P. Nyren, and M. Ronaghi. 2001. Method enabling fast partial sequencing of cDNA clones. *Anal. Biochem.* 292:266-271.
- Nygren, M., M. Ronaghi, P. Nyren, J. Albert, and J. Lundeberg. 2001. Quantification of HIV-1 using multiple quantitative polymerase chain reaction standards and bioluminescent detection. *Anal. Biochem.* 288:28-38.
- Paabo, S., J.A. Gifford, and A.C. Wilson. 1988. Mitochondrial DNA sequences from a 7000-year-old brain. *Nucleic Acids Res.* 16:9775-9787.
- Ramsay, G. 1998. DNA chips: state-of-the-art. *Nat. Biotechnol.* 16:40-44.
- Reynolds, R., G. Sensabaugh, and E. Blake. 1991. Analysis of genetic markers in forensic DNA samples using the polymerase chain reaction. *Anal. Chem.* 63:2-15.
- Reynolds, R., K. Walker, J. Varlaro, M. Allen, E. Clark, M. Alavaren, and H. Erlich. 2000. Detection of sequence variation in the HVII region of the human mitochondrial genome in 689 individuals using immobilized sequence-specific oligonucleotide probes. *J. Forensic Sci.* 45:1210-1231.
- Ronaghi, M. 2000. Improved performance of pyrosequencing using single-stranded DNA-binding protein. *Anal. Biochem.* 286:282-288.
- Ronaghi, M. 2001. Pyrosequencing sheds light on DNA sequencing. *Genome Res.* 11:3-11.
- Ronaghi, M., S. Karamohamed, B. Pettersson, M. Uhlen, and P. Nyren. 1996. Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* 242:84-89.
- Ronaghi, M., B. Pettersson, M. Uhlen, and P. Nyren. 1998. PCR-introduced loop structure as primer in DNA sequencing. *BioTechniques* 25:876-884.
- Ronaghi, M., M. Uhlen, and P. Nyren. 1998. A sequencing method based on real-time pyrophosphate. *Science* 281:363-365.
- Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5467.
- Stoneking, M., D. Hedgecock, R.G. Higuchi, L. Vigilant, and H.A. Erlich. 1991. Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *Am. J. Hum. Genet.* 48:370-382.
- van Oorschot, R.A. and M.K. Jones. 1997. DNA fingerprints from fingerprints. *Nature* 387:767.
- Vigilant, L. 1999. An evaluation of techniques for the extraction and amplification of DNA from naturally shed hairs. *Biol. Chem.* 380:1329-1331.
- Walsh, P.S., D.A. Metzger, and R. Higuchi. 1991. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques* 10:506-513.
- Wilson, M.R., D. Polansky, J. Butler, J.A. DiZinno, J. Replogle, and B. Budowle. 1995. Extraction, PCR amplification, and sequencing of mitochondrial DNA from human hair shafts. *BioTechniques* 18:662-669.

Received 21 June 2001; accepted 21 August 2001.

Address correspondence to:

Dr. Marie Allen
Uppsala University
Department of Genetics and Pathology
Rudbeck Laboratory
75185 Uppsala, Sweden
e-mail: marie.allen@genpat.uu.se