Int J Leg Med (1992) 105:83–86

International Journal of
Legal Medicine

Identification of human remains by amplification and automated sequencing of mitochondrial DNA

K. M. Sullivan, R. Hopgood, and P. Gill

Central Research and Support Establishment, The Forensic Science Service, Aldermaston, Reading, Berkshire, RG7 4PN, UK

Received December 2, 1991 / Received in revised form April 23, 1992

Summary. The highly decomposed remains of a corpse were identified by the amplification and direct sequencing of mitochondrial (mt) DNA. Degraded DNA was extracted from bone fragments and a necrotic skin sample and amplified at 2 hypervariable segments within the mitochondrial non-coding region using 2 rounds of nested PCR. Both strands of the amplified regions were sequenced and compared with each other to ensure fidelity of the data. Sequences from the bone and skin were found to be identical and matched data generated from a blood sample provided by a presumptive sister of the deceased. Thus, this evidence was consistent with the sister and the deceased being related.

Key words: Polymerase Chain Reaction – Sequencing – Mitochondrial DNA– Non-coding region – Human identification

Zusammenfassung. Die hochgradig zersetzten Überbleibsel einer Leiche wurden mit Hilfe der Amplifikation und der direkten Sequenzierung der mitochondrialen DNA identifiziert. Degradierte DNA wurde von Knochenfragmenten und von einer nekrotischen Hautprobe extrahiert und in zwei hypervariablen Segmenten der mitochondrialen nicht-kodierenden Region amplifiziert unter Anwendung zweier Runden von „nested" PCR. Beiden Stränge der amplifizierten Regionen wurden sequenziert und miteinander verglichen, um die Vertrauenswürdigkeit der Daten zu sichern. Die Sequenzen von Knochen und von der Haut waren identisch und paßten zu den Daten aus einer Blutprobe, welche von einer mutmaßlichen Schwester des Verstorbenen herrührte. Dieser Beweis war somit konsistent mit der Annahme, daß Schwester und Verstorbene verwandt waren.

Schlüsselwörter: Polymerase-Kettenreaktion – Sequenzierung – Mitochondriale DNA – Nicht-kodierende Region – Menschliche Identifizierung

Introduction

There are a number of compelling advantages to the use of mitochondrial DNA as a substrate for direct sequencing in forensic applications; the technique is extremely sensitive with between 1000 and 10000 copies of the mt genome present in each human cell [1]; mitochondria are maternally inherited in humans [2] which eliminates potential problems in interpreting overlapping heterozygous sequencing ladders; most mtDNA variation is within the non-coding region which contains the origin of replication for one strand, both origins of transcription and the D-loop region [3]. This sequence variation is specifically concentrated in 2 segments of about 400 bp each, which fall within the D-loop and around the origin of replication [4–7], and are shown in Fig. 1 as Segments A and B respectively. Each segment can be individually analysed in a single sequencing reaction. This hypervariability within the non-coding region has been utilised for human identification purposes [8, 9] and individuality can be readily determined by combining PCR amplification with direct DNA sequencing [10, 11]. Alternatively, the amplified DNA can be characterised by hybridisation with sequence-specific oligonucleotide (SSO) probes [12].

A potential drawback to mtDNA sequencing is the labour intensive nature of the technique; for routine forensic analysis the process must be highly automated to maximise sequence throughput, minimise errors in data handling, and to streamline database management [13]. This has now been achieved through recent advances in PCR strategies coupled with fluorescence-based automated sequencing technology [14].

This paper describes a case in which mitochondrial amplification and automated sequencing was used to identify highly degraded human remains, and discusses the utility of this approach as a means for routine forensic identification.

Materials and methods

Case summary. In July 1990 the body of a female, in an advanced state of decomposition, was discovered in an open field. Identification was not possible from clothing or fingerprints, but dentition was consistent with old dental records of a missing person from the area. Fragments of heel bone and fibula plus samples of head hair and skin were provided for DNA analysis, together with a liquid blood sample from a putative sister of the deceased.

DNA extraction and quantitation. DNA was extracted from the following sample: 300 ng of head hair, a 4 cm^2 piece of skin and 2

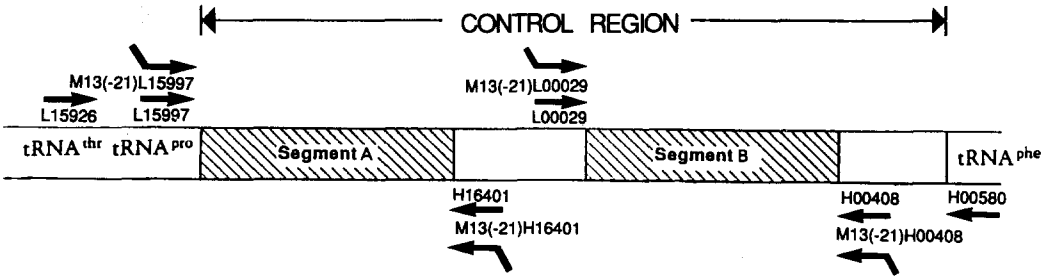

Fig. 1. Diagram of mtDNA control region. Primers used in first and second rounds of amplification are shown as arrows. Primers L15926, L15997 and H00580 fall within the threonine, proline and phenylalanine tRNA genes respectively. L15926 and H00580 were used in the first round of amplification. In the second round, 2 pairs of primers were used in separate reactions to amplify segment A: M13(−21)L15997 with H16401 and L15997 with M13(−21)H16401, and pairs to amplify segment B: M13(−21) L00029 with HJ00408 and L00029 with M13(−21) H00408

bone fragments approximately 8 mm in diameter from the corpse, plus a 500 µl blood sample from a putative sister of the deceased. Skin, bone and hair samples were powdered in liquid nitrogen using a Spex Freezer Mill. The bone sample was subsequently decalcified by soaking the fragments in 1M EDTA for 36 h [15, 16]. All samples including the liquid blood were then incubated overnight at 37°C with Proteinase K, and dithiothreitol in buffered sodium dodecyl sulphate [17]. This mixture was extracted twice with phenol and the DNA was precipitated with 2.5 vol. absolute ethanol, 0.1 vol. sodium acetate (pH 5.6) and 20 µg glycogen carrier at −80°C for 1 h. The DNA was pelleted by spinning in a microcentrifuge for 10 min and the pellet was then washed in 70% ethanol and vacuum dried briefly. The pellet was resuspended in 50 µl dH_2O, dialysed extensively against TN buffer (10 m*M* NaCl, 10 m*M* Tris/HCl pH 7.6) and quantitated fluorometrically with a Hoefer TKO 100 Mini Fluorometer with Hoechst dye 334258 (Hoefer Scientific Instruments, San Francisco, CA, USA) [18]. Human DNA content was determined by hybridisation with a human specific DNA probe kit (PhotoProbe™ Human DNA Quantitation System, Gibco BRL, New York, USA) according to the manufacturer's instructions. Briefly, this entailed immobilising aliquots of the extracted DNA on a nylon membrane (BioDyne A, Pall Europe Limited, Portsmouth, England) by means of a slot-blot apparatus (Convertible™ Filtration Manifold System, Gibco BRL, New York, USA) connected to a vacuum source. The DNA was then hybridised with an alkaline phosphatase-labelled oligonucleotide (D17Z1), and after the addition of Lumi-Phos 530™ (Lumigen Inc.) to the membrane, the hybridised probe was visualised using photographic film.

DNA amplification. Primers used in amplification were as follows:
L15997 (5′-CACCATTAGCACCCAAAGCT) [9],
H16401 (5′-TGATTTCACGGAGGATGGTG) [9],
L15926 (5′-TCAAAGCTTACACCAGTCTTGTCTTGTAAACC) [9],
H00580 (5′-TTGAGGAGGTAAGCTACATA) [9],
M13(−21)L15997 (5′-TGTAAAACGACGACGGCCAGTCACCATTAGCACCCAAAGCT) [12],
M13(−21)H16401 (5′-TGTAAAACGACGACGGCCAGTTGATTTCACGGAGGATGGTG) [12],
L00029 (GGTCTATCACCCTATTAACCAC) [10],
H00408 (CTGTTAAAAGTGCATACCGCCA) [10],
M13(−21)L00029 (TGTAAAACGACGACGGCCAGTGGTCTATCACCCTATTAACCAC)
M13(−21)H00408 (TGTAAAACGACGACGGCCAGTCTGTTAAAAGTGCATACCGCCA)

The latter 4 primers are chimeric, comprising a mitochondrial PCR primer sequence at the 3′ end and the M13 universal sequencing primer sequence at the 5′ end. Amplficiation with one of these primers incorporates the universal primer sequence in the PCR product so that corresponding primers can be utilised in the subsequent sequencing reactions. It should be noted that these primers incorrectly included an additional 3 nucleotides CGA at positions 8–10 from the 5′ end, compared with M13(−21) universal sequencing primer sequence, but this does not appear to affect the reactions. All amplification reactions were performed using a Perkin Elmer Cetus 480 thermal cycler.

In the first round of PCR the entire mitochondrial non-coding region was amplified (Fig. 1): 20–200 ng of total extracted DNA was subjected to 30 rounds of amplification in a 25 µl reaction volume containing 1.5 U Taq polymerase, 200 µ*M* each dNTP, 400 ng/µl BSA, 10% Triton X100 and 1 µ*M* primers L15926 and H00580. Each cycle comprised 45 seconds at 94°C, 1 min at 50°C and 5.5 min at 72°C.

Amplification products were monitored by electrophoresing 10 µl aliquots in a 1% agarose gel for 2 h at 100 V. Aliquots of PCR mixture 0.5 µl containing the resultant 1333 bp amplification product were added to an array of second round amplification reactions to generate sufficient DNA template for sequencing, using the following conditions: 45 s at 94°C, 1 min at 50°C and 3 min at 72°C for 32 cycles in a 50 µl reaction containing 2.5 U Taq polymerase, 100 ng/µl BSA, and 20 µM each dNTP. Where the yield of single-stranded product was low, amplification was repeated using 200 µM each dNTP which generally improved the ratio of single-stranded to double-stranded product. To specifically amplify hypervariable segment A, primers L15997 and M13(−21)H16401 were used in unequal concentrations (1 µM: 20 n*M*) thereby generating single-stranded "L-strand" DNA for direct sequencing. In a separate reaction the "H-strand" of segment A was preferentially amplified with primers H16401 and M13(−21)L15997 (1 µ*M*: 20 n*M*). Similarly hypervariable segment B was preferentially amplified in 2 asymmetric PCR reactions using the primer pairs L00029 with M13(−21)H00408, and H00408 with M13(−21)L00029 at concentration ratios of 1 µ*M*:20 n*M* respectively. Amplification products were verified on ethidium bromide-stained (0.5 µg/ml) 3% Nusieve +1% Seaplaque (FMC BioProducts Rockland, ME, USA) agarose gels: 15 µl aliquots were electrophoresed for 45 min at 100 V. Short electrophoretic times coupled with a high percentage gel concentration enabled both the double-stranded band and the more diffuse single-stranded PCR product to be readily visualised.

Sequencing reactions. For the dideoxyadenosine sequencing reactions, 1.5 µl of the PCR solution was added to the following: 0.8 pmol of −21 M13 "JOE" primer (Applied Biosystems); 2.5 U Taq polymerase; 5 × sequencing buffer to give a final concentration of 10 m*M* Tris-HCl, pH8.0, 10 m*M* $MgCl_2$, 25 m*M* NaCl; 2 µl of 2′deoxy/2′,3′-dideoxyadenosine-5′triphosphate mix; dH_2O to a final volume of 15 µl. The three other sequencing reactions for dideoxycytosine, dideoxyguanosine and dideoxythymidine were set up in parallel using "FAM", "TAMRA" and "ROX"-labelled primers respectively, with the same conditions except that all quantities were doubled for the dideoxyguanosine and dideoxy-

thymidine reactions as detailed in the ABI 370A Sequencer users manual (ABI, Foster City, California, USA). These mixtures were heated as follows: 90°C, 1 min; 70°C, 2 min; for 10 cycles. The reactions were then pooled and precipitated with ethanol, resuspended in 5 µl deionized formamide and 1 µl 50 m*M* EDTA pH 8.0, denatured for 2 min at 95°C and loaded directly on to a 6% polyacrylamide gel. Electrophoresis and sequence analysis were performed with an Applied Biosystems Model 370A DNA Sequencer. Sequence comparisons were performed using a DNASTAR (Dnastar Ltd, London, UK) analysis package.

Results

Yields of extracted DNA were generally small and of low molecular weight. Fluorometric readings indicated that the total nucleic acid contents of the samples were as follows: 10 µg per gram of bone, 1 µg per cm^2 skin, and 0.5 µg per mg head hair. In contrast, quantitation by hybridisation assay with a human chromosomal DNA probe indicated that approximately 50 ng DNA per gram bone, 3 ng per cm^2 skin and < 200 pg per mg head hair were of human origin. Thus, it is likely that most of the isolated DNA was of microbiological origin [15].

All samples yielded amplification products after 2 rounds of PCR (Fig. 2). However, the yield of single-stranded DNA was variable when using 20 µ*M* of each dNTP in the second PCR reaction. This was circumvented by re-amplifying with a higher dNTP concentration (200 µ*M*) then removing excess unincorporated nucleotides by ethanol precipitation of the DNA followed by 2 washes in 70% ethanol, prior to sequencing. This purification step removes excess nucleotides which would otherwise interfere with the deoxy: dideoxynucleotide ratio of the sequencing reactions resulting in more unresolved bases.

The DNA sequences of hypervariable segments A and B were obtained from the bone, skin and blood DNA samples. A control DNA sample from the PCR operator was characterised for segment A. Full length sequencing ladders were obtained from all samples with *c.*350 bp generated from a typical optimised reaction. When using dye-labelled primers and Taq cycle sequencing to generate extension products, the length of readable sequence is influenced by the size of the peaks in the raw data corresponding to the primer and full-extension products. Raw data between these peaks is selected for subsequent sequence analysis. In our experience the size of these two raw data peaks is sample-dependent, thus the total sequence length generated from different samples also varies, and is the subject of further investigation into the optimised sequencing of PCR products [14]. Both L and H strands were sequenced at each hypervariable segment and sequence complementarity was checked by computer to generate a consensus sequence for each sample. These were then used in comparisons with other sequence data. No differences were observed between the skin and bone DNA of the corpse or between blood from the putative sister and bone from the corpse. In contrast, bone from the corpse and control DNA from the operator differed by 3 nucleotides at position 16224 (C and T respectively), 16304 (T and C) and 16311 (C and T), using the nucleotide numbering system of Anderson et al. [3]. These results are summarised in Table 1.

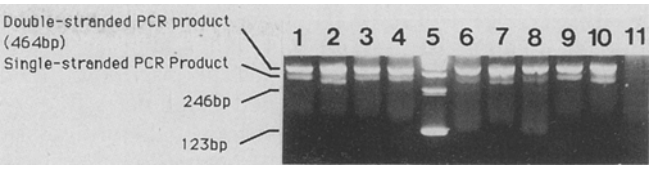

Fig. 2. Second round amplification products electrophoresed in a 3% Nusieve + 1% Seaplaque gel. Lanes 1 and 2: blood from putative sister amplified using primer pairs L15997 with M13(−21) H16401 and M13(−21)L15997 with H16401 respectively; lanes 3 and 4: DNA extracts from skin both amplified with L15997 and M13(−21)H16401; lane 5: 123 bp standard; lanes 6 and 7: bone DNA using L15997 with M13(−21)H16401 and M13(−21)L15997 with H16401 respectively; lane 8: DNA extracted from hair amplified with L15997 and M13(−21)H16401; 9 and 10: operator control DNA using L15997 with M13(−21)H16401 and M13(−21) L15997 with H16401 respectively; lane 11: negative control with L15997 and M13(−21)H16401

The shared corpse/putative sister sequence of segment A was compared against 130 published Caucasian sequences of worldwide origin [4, 6, 9, 11, 19]: 3 out of 130 matched (data not shown). Of these 3 DNA samples that matched, only 1 had also been sequenced within segment B and was different to the corpse/putative sister sequence.

Discussion

This study demonstrated the utility of mitochondrial sequencing as a method for human individualisation in forensic investigations. Results can be obtained from trace amounts of highly degraded DNA, plus the sequence data is definitive and can be highly informative. In this investigation, a statistical value could not be placed on the evidence since a database of British Caucasian sequences was not available at the time. However, no differences were found between sequences from the blood of the putative sister and bone of the corpse, which was consistent with the two being sisters. Approximately 3% (3 out of 130) of the published Caucasian sequences for segment A matched this shared sequence, but little sequence data is available for segment B. A preliminary study of 14 Caucasians by other workers indicated that pairwise comparisons of segment A sequences gave a Poisson distribution for the number of nucleotide differences between individuals. Thus, the probability of a random match (pM) between pairs of individuals was determined to be 1 in 370 from the equation $pM = e^{-n}$, where n = mean difference of 5.9 [10]. However, this calculation was based on the premise that all observed differences are statistically independent, which is not necessarily correct. In another study, 142 US Caucasians were characterised by hybridisation assay using SSO probes to relevant sequences within both hypervariable segments [12]. This established an overall probability of a random match (pM) of 0.019 on the

basis of observed mtDNA types. More sequence analysis will be required to determine whether an appropriate statistical model can be applied to the population data rather than simply quoting observed frequencies. We are currently addressing this problem in our laboratory by constructing a British Caucasian database for both hypervariable regions.

As with all PCR techniques, contamination is a major concern in direct mitochondrial sequencing due to the large number of amplification cycles involved. Suitable controls should be used such as typing water blanks which are taken through the entire extraction process, typing of laboratory operators, and ideally sequencing more than 1 sample from an individual. The latter precaution also acts as a check for sequencing accuracy: a 1% base-calling error is typical when analysing a single strand of a 400 bp segment by automated sequencing. Thus, characterisation of the complementary strand should reduce this error rate to 0.01%, and comparison of duplicate samples further reduces potential errors. Any discrepancies between L and H strand sequences were resolved by manual comparison of the sequence electrophoretograms. Fortunately all discrepancies encountered with this case comprised a resolved base on 1 strand versus an unresolved base (scored by computer as "N") on the complemetary strand. However, situations could arise where manual comparison does not resolve the base call such as when a particular segment of the DNA yields poor sequence from both strands. Under these circumstances, the position would be recorded as "N" in the consensus sequence and therefore would serve no part in subsequent sequence comparisons.

A major concern in the use of mtDNA sequences for forensic purposes is potential heterogeneity within an individual. However, a high degree of sequence homogeneity has been demonstrated in somatic cells [20, 21], and although heteroplasmy (i.e. a mixture of mutant and wildtype mtDNAs in the same cell) has been reported in humans this only appears to be associated with certain neuromuscular diseases and has not been observed in the non-coding region [12, 22]. Further studies of families are required to determine the rate at which non-coding nucleotide mutations become fixed within a given maternal lineage. Optimisation of sequence analysis and processing both strands of a given DNA segment will enable true sequence microheterogeneities to be distinguished from sequence artefacts such as extension "pause" sites. A further potential source of sequence microheterogeneity is paternal inheritance of mtDNA. This was recently detected in mice at a frequency of 10^{-4} relative to the maternal contributions [23], and this maternal/paternal sequence heteroplasmy could also exist in humans. However, provided several mtDNA molecules from a given individual are amplified in the same reaction, any potential paternal sequence contribution will remain undetectable as background noise. Despite these concerns, mtDNA sequencing displays great promise for routine forensic investigations, particularly in the light of recent advances in PCR techniques and automated sequencing technology.

References

1. Bogenhagen D, Clayton DA (1974) The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. J Biol Chem 249:7991–7995
2. Giles RE, Blanc H, Cann HM, Wallace DC (1980) Maternal inheritance of human mitochondrial DNA. Proc Natl Acad Sci USA 77:6715–6719
3. Anderson S, Bankier AT, Barrell BG, deBruijin MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organisation of the human mitochondrial genome. Nature 290:457–465
4. Aquadro CF, Greenberg BD (1983) Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. Genetics 103:287–312
5. Greenberg BD, Newbold JE, Sugino A (1983) Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. Gene 21:33–49
6. Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325:457–465
7. Horai S, Hayasaki K (1990) Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. Am J Hum Genet 46:828–842
8. Higuchi R, von Beroldingen CH, Sensabaugh GF, Erlich HA (1988) DNA typing from single hairs. Nature 332:543–546
9. Reynolds R, Sensabaugh G, Blake E (1991) Analysis of genetic markers in forensic DNA samples using the polymerase chain reaction. Anal Chem 63:2–15
10. Orrego C, King MC (1990) Determination of familial relationships. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) PCR Protocols, Academic Press, London, 416–426
11. Vigilant L, Pennington R, Harpending H, Kocher TD (1989) Mitochondrial DNA sequences in single hairs from a southern African population. Proc Natl Acad Sci USA 86:9350–9354
12. Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA (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
13. Sullivan KM, Hopgood R, Lang B, Gill P (1991) Automated amplification and sequencing of human mitochondrial DNA. Electrophoresis 12:17–21
14. Hopgood R, Sullivan KM, Gill P (1992) Strategies for automated sequencing of human mitochondrial DNA directly from PCR products. Biotechniques (in press)
15. Hagelberg E, Gray IC, Jeffreys AJ (1991) Identification of the skeletal remains of a murder victim. Nature 352:427–429
16. Hochmeister MN, Budowle B, Borer UV, Eggman U, Comey CT, Dirnhofer R (1991) Typing of deoxyribonucleic acid (DNA) extracted from compact bone from human remains. J Forensic Sci 36:1649–1661
17. Gill P, Lygo JE, Fowler SJ, Werrett DJ (1987) An evaluation of DNA fingerprinting for forensic purposes. Electrophoresis 8:35–38
18. Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor New York
19. Di Renzo A, Wilson AC (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc Natl Acad Sci USA 88:1597–1601
20. Monnat RJ, Loeb LA (1985) Nucleotide sequence preservation of human mitochondrial DNA. Proc Natl Acad Sci USA 82:2895–2899
21. Monnat RJ, Reay DT (1986) Nucleotide sequence identity of mitochondrial DNA from different human tissues. Gene 43:205–211
22. Wallace DC (1991) Mitochondrial DNA mutations and neuromuscular disease. Trends Genet 5:9–13
23. Gyllensten U, Wharton D, Josefsson A, Wilson AC (1991) Paternal inheritance of mitochondrial DNA in mice. Nature 352:255–257