# ORIGINAL ARTICLE

N. Dimo-Simonin · F. Grange · F. Taroni C. Brandt-Casadevall · P. Mangin

# Forensic evaluation of mtDNA in a population from south west Switzerland

Received: 19 January 1999 / Accepted: 5 May 1999

**Abstract** The polymorphism of the two hypervariable segments (HVI and HVII) of the control region of mtDNA was analyzed in a population of 154 unrelated individuals from south west Switzerland using a fluorescent based capillary electrophoresis sequencing method.

In our population data of 154 random individuals, 137 mtDNA types were observed. Of these, 124 sequences were observed only in one individual whereas 10 sequences were observed in 2 individuals, 2 sequences in 3 individuals and 1 sequence in 4 individuals. The probability of two unrelated individuals having the same sequence was 0.84%. The results were compared with four other Caucasian populations.

Furthermore, the usefulness of the mtDNA sequencing was tested, for exclusion and inclusion, in 18 forensic cases including 69 evidence samples and 44 reference samples. Despite the fact that 55% of the evidence samples yielded a negative result for the nuclear DNA with the human dot quantitation system, the success rate of the mtDNA sequencing was 71.0%. This validation study proves the great usefulness and sensitivity of the mtDNA sequencing technique using nested PCR and fluorescent capillary electrophoresis.

Key words Mitochondrial DNA  $\cdot$  Control region  $\cdot$  PCR  $\cdot$  Sequencing  $\cdot$  Population study  $\cdot$  Forensic casework

# Introduction

During the last 10 years, the PCR-based DNA systems have been a powerful tool for the forensic identification of human biological evidence. Despite the great usefulness and sensitivity of the polymorphic nuclear DNA loci for the

C. Brandt-Casadevall · P. Mangin

Institut Universitaire de Médecine Légale, Rue du Bugnon 21, Lausanne, Switzerland forensic laboratory, some casework samples with highly degraded or very little DNA may give inconclusive results. In these cases, the analysis of mitochondrial DNA (mtDNA) can be successful. Furthermore, automated fluorescence based electrophoresis for sequence analysis of PCR samples has allowed the use of the sequencing analysis of mtDNA in routine casework.

It was soon discovered that the mtDNA is highly polymorphic and present in multi copy numbers per cell [1]. In addition, the molecule is small, involving a double-stranded closed circular structure of 16569 bp [2] and is maternally inherited [3]. These characteristics are forensically significant and provide the basis for the identification of hair shafts, degraded bones and samples with very little or degraded DNA.

Recently, the use of mtDNA sequencing has gained increasing popularity in the forensic community and several studies concerning casework [4–12] and population data [13–19] have been reported.

The frequency of mtDNA sequence in a given population being based on the counting method, the aim of this study was to determine the polymorphism of the two hypervariable segments (HVI and HVII) of the non–coding control region in a South West Swiss population of 154 unrelated french speaking individuals using a nested PCR amplification with fluorescent based capillary electrophoresis sequencing method. Furthermore, comparison with other Caucasian populations [13–18] and usefulness of this technique for exclusion and inclusion in some forensic casework is reported.

### **Materials and methods**

Extraction

Population data

Blood samples from 154 unrelated individuals from the Lausanne, Fribourg and Sion areas (South-western Switzerland) were saline extracted and the nuclear DNA concentration was tested at 260 nm with a spectrophotometer.

N. Dimo-Simonin (⊠) · F. Grange · F. Taroni

#### Casework samples

Samples from 18 forensic cases including 15 bloodstains, 4 fingernails, 4 bones, 1 muscle sample, 22 hairs, 16 sweat stains, 6 saliva stains, 1 serum and 44 blood and saliva reference samples were typed. The hairs segments were washed with sterile water and ethanol (100%) before the extraction.

The muscle and bone samples were extracted using the phenolchloroform method followed by the chelex method. The other samples were chelex [20] extracted with some minor modifications. All the DNA extracts were further purified and concentrated by membrane-based size exclusion using a Centricon 100 spin column (Amicon USA). A quantification of the DNA extracts by a dot method with the human DNA quantitation System (Life Technologies USA) was then performed.

#### Amplification

A range of < 0.3 up to 1 ng of DNA was used for the nested PCR amplification [21]. In the first round PCR, the complete 1333 bp mitochondrial non-coding control region was amplified using the primers L15926 and H00580 (Table 1). For the population data, template DNA was subjected to 30 cycles of amplification in 25 µl reaction volume containing 50 mM KCl, 10 mM Tris-HCl (pH 9.0 at 25 °C), 1% Triton X-100, 1.5 mM MgCl<sub>2</sub>, 0.1 mM each dNTP, 0.5 µM each primer, 1.5 U Taq Polymerase (Promega USA), 10 µg BSA and 50 µl mineral oil. The forensic samples were amplified with the Expand high Fidelity PCR system (Boehringer Mannheim, Germany) using the amplification mix with 1.5 mM MgCl<sub>2</sub>, 0.1 mM each dNTP, 0.5 µM each primer, 1.5 U high Fidelity polymerase and 50 µl mineral oil. Amplification was performed in the Biometra triothermoblock at 94 °C for 45 s, 50 °C for 60 s and 72 °C for 5 min 30 s for 30 cycles. The size and quantity of the PCR products were estimated by UV fluorescence electrophoresis in 2% agarose gels containing ethidium bromide for 30 min at 100 V using the Low DNA Mass Ladder (Life Technologies USA).

Aliquots of PCR mixture  $(0.5-4 \ \mu)$  were added to a second PCR reaction using primers M13 (-21) L15997 and M13 Rev H16401 for HVI and M13 Rev L00029 and M13 (-21) H00408 for HVII. The primers used for this second amplification were chimeric or tailed primers [4, 15] that consist of universal M13 (M13-21) or M13 reverse (M13 Rev) tails at the 5' end plus the gene specific sequence (Table 1). The nested amplification was performed for 30 cycles in a 25  $\mu$ l reaction volume using the same mix as described except that 0.5 U Taq polymerase (Promega and Expand high Fidelity Taq polymerase for the population and the forensic samples, respectively) and 2  $\mu$ g BSA were used. The amplification conditions were 94 °C for 45 s, 60 °C for 60 s and 72 °C for 3 min.

Table 1         Sequence of the sequence of th	the H	ivi an	ануп	primers
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------	--------	------	---------

Primer	Sequence
L15926	5'-TCA AAG CTT ACA CCA GTC TTG TAA ACC-3'
H00580	5'-TTG AGG AGG TAA GCT ACA TA-3'
M13(-21) L15997	5'-TGT AAA ACG ACG GCC AGT CAC CAT TAG CAC CCA AAG CT-3'
M13 Rev H16401	5'-CAG GAA ACA GCT ATG ACC TGA TTT CAC GGA GGA TGG TG-3'
M13 Rev L00029	5'-CAG GAA ACA GCT ATG ACC GGT CTA TCA CCC TAT TAA CCA C-3'
M13(-21) H0048	5'-TGT AAA ACG ACG GCC AGT CTG TTA AAA GTG CAT ACC GCC A-3'
L1597	5'-CAC CAT TAG CAC CCA AAG CT-3'
H16401	5'-TGA TTT CAC GGA GGA TGG TG-3'
L00029	5'-GGT CTA TCA CCC TAT TAA CCA C-3'
H00408	5'-CTG TTA AAA GTG CAT ACC GCC A-3'

The DNA concentrations of HVI and HVII amplicons were estimated as described for the 1333 bp fragment and then purified with a QIAquick spin PCR purification Kit (Qiagen, Germany) to remove remaining amplification primers, dNTP and buffer. The elution was done with 30–100  $\mu$ l elution buffer according to the HVI or HVII nested PCR product quantity estimated by agarose gel electrophoresis. Purified DNA (1  $\mu$ l) was estimated by agarose minigel electrophoresis to determine the optimal PCR product quantity for the sequencing analysis.

#### Cycle sequencing and capillary electrophoresis

Sequencing reactions were performed in a Biometra triothermoblock or a thermoblock Progene (Techne) using the ABI (Applied Biosystems) Prism dRhodamine or BigDye Terminator Cycle Sequencing Ready Reaction Kits with AmpliTaq DNA polymerase FS (Perkin Elmer). For HVI (forward and reverse reaction) and HVII (forward reaction), about 3-10 ng of purified amplicon was used as template in 10 µl (one half the normal reaction volume and kit reagents) for 25 cycle at 96 °C for 30 s, 55 °C for 15 s and 60 °C for 4 min. The HVII reverse reaction was done with the same protocol but at 60 °C annealing temperature. Both strands were sequenced using 3.2 pmol of each primer: L15997 and H16401 for HVI and L00029 and H 00408 for HVII (Table 1). The sequencing products were then purified using the Centri-sep spin columns (Princeton Separation USA), prepared and analyzed on a ABI 310 Genetic Analyzer with the ABI Prism Sequencing Analysis Software version 3.0 according to the ABI protocol (Applied Biosystems USA). The sequences were then aligned and analyzed with the Factura and Sequence Navigator softwares versions 1.2 and 1.02b3 respectively. The denatured samples were electrokinetically injected at 2.0 kV for 30 s into a 47 cm uncoated capillary. The electrophoresis was run at 15.0 kV and 50 °C for 36 min with the POP 6 polymer and the rapid sequencing module.

#### Statistical analysis

The mean nucleotide pairwise difference was calculated using an unpublished program by Dr. G. van Melle (Institut Universitaire de Médecine Sociale et Préventive, University of Lausanne). This estimation was done for five Caucasian populations considering the HVI (16030–16400) and HVII (40–390) regions: French [18], German [15, 16], Austrian [17], British [13] and South-West Swiss (this paper). Furthermore, the matching probability was also calculated according to the formula:

matching probability 
$$=\sum_{i=1}^{h} (x_i/N_x)$$

where  $x_i$  is the absolute frequency of the *i*-th haplotype and h the total number of different haplotypes in the  $N_x$  samples. For small sample sizes, the formula is multiplied by n/(n-1).

## **Results and discussion**

Nested amplification and cycle sequencing

Half of our forensic evidence samples yielded a very low nuclear DNA quantity and the majority did not yield a visible PCR product after the first amplification (Table 2). In these instances, 4  $\mu$ l of the first amplicon was added to the second PCR reaction and 81.2% of the evidence samples generated sufficient PCR product to be sequenced several times. The Expand high Fidelity System gave an optimal nested PCR amplification and seemed to increase the 1333 bp PCR product in comparison with other poly-

Table 2 mtDNA nested PCR success rates of the forensic evidence samples (n = 69)

Results	HVI + HVII 1333 bp	HVI and HVII 400 bp
Negative	40.6%	15.9%
Slightly positive	17.4%	2.9%
Positive	42.0%	81.2%

 Table 3
 HVI and HVII heavy related sequencing problems

	5 1	01
Specific sites	HVI	HVII
Compression	16351/16352 (T/A) 16359/16360 (A/G) 16368/16369 (A/C) 16372/16373 (A/C)	321/322 (A/C) 341/342 (T/A) 372/373 (A/T) 383/384 (A/T) 390/391 (T/A)
Green artefacts peaks	16262* 16287* 16296* 16328** 16355*	314*** 317* 320* 345* 371*

merases (manuscript in preparation). In agreement with some authors [4, 6, 10, 21], the nested PCR seemed to be optimal to obtain good sequencing results in cases with very low DNA quantity. The disadvantage of this method is the high sensitivity to contamination from other samples [4, 10] but this can be monitored with extraction and PCR controls. We observed a removal in contamination problem using mineral oil for the two-stage PCR reaction. In addition, samples with low or degraded DNA have to be extracted and amplified separately from the samples containing a high DNA content. It is also very important to use laminar flow hoods with dedicated reagents strictly for mtDNA analysis and to UV irradiate the extraction and amplification area and equipment some hours before utilisation [6–8, 25, 26].

## Sequencing data

The two hypervariable sequences HVI and HVII of the mtDNA control region from 154 South West Swiss unrelated individuals and 113 casework samples were tested by cycle sequencing with the Dye Terminator method. The L-strand (forward reaction) and the H-strand (reverse reaction) were sequenced with specific primers lacking a universal tail. The first advantage of using chimeric or tailed primers in the second PCR reaction was to carry out the nested sequencing method which yielded good results data from direct PCR sequencing as the primers anneal internally to the primer used for the second PCR. This eliminates the possible primer oligomers formed during the nested PCR and not totally eliminated with the QIAGEN PCR purification and the signal to noise ratio is often improved [22-24]. A second advantage was the possibility to cycle sequence the same nested PCR product with another procedure. The flexibility in using dye terminator or dye primer labeling can be useful to resolve some difficulties (ambiguity, heteroplasmy, mixture) because each sequencing procedure has particular features and the interpretation of some difficult cases can be facilitated. Furthermore, the possibility to use a universal primer instead of the specific primer with the dye terminator method may sometimes also improve the sequencing results [22–24].

The automated base calling was confirmed when the two strand sequences were identical and the mtDNA polymorphism were reported after comparison with a human MtDNA reference sequence (Anderson et al. 1981). For HVI, the differences between 16020 and 16400 and for HVII between 40 and 400 were reported.

\* small G peak after A peak

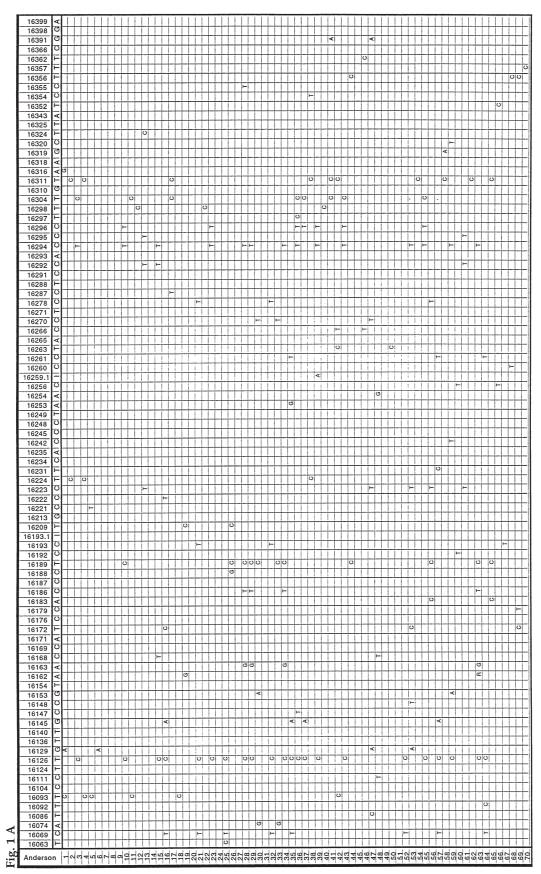
\*\* small G peak after C peak

\*\*\* small G peak in G stretch

In agreement with some authors [17], we observed that forward sequencing results were often better than reverse sequencing. It is known that the purine rich H-strand frequently gives more sequencing problems than the pyrimidine rich L-strand. This problem was visualised at different specific positions (Table 3) by sequence compression or by a green artefact peak particularly in some small G signals after an A signal, sometimes giving a false heterozygote call whereas in the opposite strand, the base call was accurate. The beginning of the sequences were the position most prone to sequencing artefacts. This problem, caused by the sequencing chemistry, can be resolved by using about 5 ng of PCR product for the cycle sequencing reaction. Using these conditions, the compression and fluorescence artefacts were decreased and the bases were reliably called by the automated software. It seemed to be very important to quantitate the PCR products for the cycle sequencing particularly with the new Big Dye procedure [27] which is very sensitive to residual dye molecules. After some preliminary assays, we modified our sequencing protocol using the 5X sequencing buffer (Applied Biosystems) with a 20 µl total sequencing volume. A decrease in artefact fluorescence, probably due to the increased efficiency of excess dye removal, was observed. At present we are testing the HVI nested PCR and the reverse sequencing reactions using a higher annealing temperature with the aim of obtaining a reduction of the slightly higher background of the HVI reverse sequencing reaction compared to the HVII reaction.

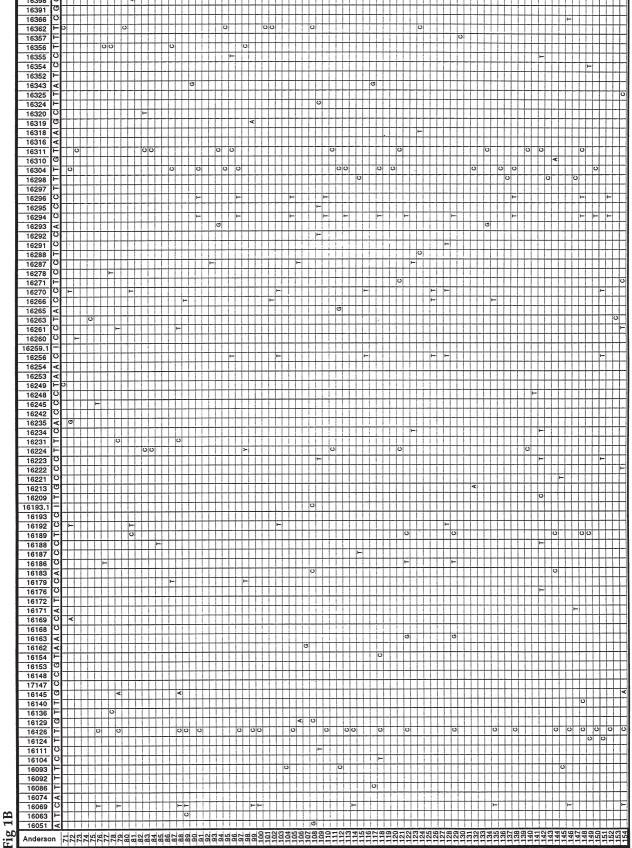
Any differences between L and H strand sequences were resolved by manual comparison of the sequence electropherograms. If an ambiguity was not resolved, the result would be "N" in the consensus sequence. In two blood samples from the population study (samples 63 and 98 in Fig. 1 A and 1B) and in two saliva samples of our casework reference material we observed a mixture of two nucleotides at positions 16093, 16162, 16224 and 150. We considered these positions as heteroplasmic as the height of the secondary peak was 50% or more of the main peak. These heteroplasmic sites were detected automatically in

382\*



**Fig.1 A–C** Variable nucleotide positions of the mtDNA D-loop regions HVI and HVII of 154 South West Swiss Caucasian samples compared to the reference sequence of Anderson et al. (1981). The differences from the reference sequence are indicated by let-

ters for base substitutions and by dashes for base insertions. N indicates no specific nucleotide. Numbers (1-3) following the integer numbers indicate addition of nucleotides not found in the reference



Anderson	00064	00072	00073	00093	00114	00119	00143	00146	00150	00151	00150	00174.1	00182	00185	00186	00188	00189	00194	16100	00100	00109	00200	00204	00207	00210	00215	00217	00225	00226	00227	00228	00239	00247	00250	00263	00279	00295	00309.1	00309.2	00309.3	00310	00311	00315.1	00315.2	00318	00319	20000	00240	00357
1	с	T	A	_			G -		_					G	-			0 1	τİΑ	÷		ΓļΑ	Ļ		ЗA	-	-	-	-	A	G	τİ		з <sup>і</sup> т	-	<sub>T</sub>	c	1		1	тİ	c	L C C			тļ		1	G
2 3 4 5	E		GI		뷲		+			+	c	+			+		G			+	+	+	+	Ť	+	+					+			+				000			+	$\pm$		q	+	+		+	_
6 7 8	E				Ì														c			-	$\frac{1}{1}$	+	-									+	GGJG			000				$\frac{1}{1}$	C C L			-		+	
9 10 11			G					C			c						a				+	+	+		1					+				1	+ G G		1	_C				$\pm$	000000000000000000000000000000000000000	d		+		+	
12		4		G	+												G		4	+	+		+	+	A	+	-			+	+	+	+	Ť	GJG	++		_C	C		-	$\pm$	C	+	+			+	_
14 15 16 17					ļ		+	c		-		+	1				+	+		i i	+	+	+	t	1	-	1			+			r	+	GGGG		T	8	c		-	+	CI CI	+	+	+	+	+	
19 20 21	E	1 1	GI		Ì			+	T	+		+	F				+		T	+		1	0	-	+		1					1	+		G	+	TI	C			-	+	CCC		+	+	+	1	
23	E	ι υ	G		Ì		Ť	Ť	1	Ť		+	-		Ŧ			+	1	T	+	1	1	T	1	-	1					1		÷	GGG	++		- 4				÷	UCCC C		-	+	1	Ŧ	_
25	I	- 4	1				+		т			+	-				+		1	-	+	+	+	÷	+	+				+	A	+	+		GGGG	1 1	+	C	C				C1	+	+	-+	+	+	_
25 26 29 30 31 32 33 34 35 35 37			G G		$\frac{1}{1}$			+	T			-	-				+		c	1	+	+	+	+	+								r!		GGG			00			1	$\pm$	0000000000000000	+	+	+	+	+	
31 32 33	E	1	G	-+			+	+	Ŧ	$\mp$		+	+				G			+	Ť		1	Ť	+	-					-		T.	+	3GGGGGGGG	$\overline{+}$		C	1 1			Ì	CCC	+	+	Ť		÷	
35	E	<u>†</u> †	GGJG	-	Ì	+	+	Ŧ	T	+	č	+	-		Ŧ		+	+				Ŧ	+	T		G									G		Ţ	coo	i i				C C	+	+	c	+	+	_
38		10	Gİ		+					+	C	1	+		-		+			1	TI	+	+	T			T T					+		-	G1GGGGJG			C				+	000	+		+			_
40 41 42 43			G		Ŧ		+	T		T	T	1						T	+	T	Ť	+	+				T								∣ G	1 1		c	1.1			+	CC				T		
44 45 46 47 48			+		-			-						H			+	+	+	+	+		-	+		-	-			_		C C	+		G			000	c	_N		+			+	-	+	+	_
47			G		+			-		+		+	+		+		Ŧ	1	+	+	+	q			+		1			-		+	Ť		GGG						+	Ì	CC	+	+		+	İ	
49 50 51 52 53 54 55 56 57	E		G		1	-	+	+		-		+	1	A		-	+	-	-	+		d	1	T	+	1					A						TI	C C	1.1			+	C			+	+	+	_
54 55 56	E	++	12		Ì		+	Ť		Ť	+	G	-				+	-	c	+		1	Ŧ	Ŧ		-	1	A	C				-	-	GGGGGG			C			1	+	C	+	+			1	
57	E	1 1	G	1				c	T			Ť					+		C	1				T		G	-			-		+	Ì	1	GGJG	<del>i i</del>	T	T U			+	-+	CI			c	+		
58 59 60 61 62 63 63 65		$\frac{1}{1}$	GGG			C		+	+	+		+	T				G	t .	c		+	+	+	+	A G	-	1			+			+		+ G	+-+						$\pm$	<u>nrnnnnnnnnnnnnn</u>	+	+	+	+	+	
63 54 65	E		G									+	1	A		_	+		c c	+	+	+	+	1	+	-					A		+		GGGGGGG		+	C	i i		$\frac{1}{1}$	$\pm$	C	+		+	+	+	
6 6 6 7 6 8 6 9			G G		$\pm$		+			+	+		1				1	1	c	Ť		Ŧ		+	+	T				Ì	Ì	+	-	1	GGG			000	с		+	$\frac{1}{1}$	C C	-	Ť	÷	Ŧ	Ţ	
70 71 72	E				+	1	-			Ì	+	+	1				-	+	+	+			+	+	+	1	-			G	A	C	+		GGG			CC			+	$\overline{+}$	CCC	+			+	+	_
74 75	E	Ħ			+		ļ			-	Ŧ	+		H		-	G	+	Ŧ	+		+	+	+		-	-					+	+	+	0000000000			с С				$\mp$	0000		+		+	-	_
76.			GG		-									A					C!		Ţ	¢	+	+							A		+		GGGGG		다	-C				$\pm$	00000		+			+	
79 80 81	-	+	G					1		+			+				+	+	c	T	+	+		+		G	-			+	1	c	+	+	G		-+	С	C		+	$\pm$	C	1	+		+		
82 83 84	E		GG		+		+	c c				+					+	+	-	+		+	+	+	-	+						-	+	+	0000000			С			+	$\frac{1}{1}$	CCCCC	+	+		+	+	
85 86 87 88	E	11			+				-		c	+	+				+	Ť	C I	+	+	+	İ	Ť	-	G	-			-		-	+		G	1	T		C						T	c	+	Ť	
90		++	GGG	-				+	T	1		1	-		_		+	+	1	T	-		1	T		Ŧ		1			A	+	+	-	G		T	C C				+			T	+	+		_
91 92 93 94	E		G	Ì	-			c										-	c			Ť	+	+		1						+	+	-	G	H											-	+	_
95 96 97	-		G							1	C		-				+	-	-	1	+	+	+	+		-	-	+				+	+					C C	1 1		+	$\pm$	0000		+		+		
98 99 100			G		+			$\frac{1}{1}$		+		-	+	A			+	+	C	+	+	+	+			+	1				A				GGGG		Ŧ	C		_			C C C	+	+		+	+	
102 103 104	E		G					+			-		1									+		+	+							C			- (G		1	C C C C				-	C		+	+			_
105	E	÷÷	G					4		+	+	+	T	T					C	+	+	+		2	+	1	-						T	A	GGGG			Ç				÷	0000					÷	G
108 109 110		1 1	GGGG					+	T		c	+	+				G	+	c	+			+	+	A		C						1	_	GGG	1.1		C				$\pm$	0000			+	+		
$\frac{112}{113}$		++	G	1			_			-	c	+	1			_		+	+	+	Ť	+	+	Ŧ	+		<u> </u>				_		Ţ	-	GGG	1 1	T	C C	C			+	C C C	+	1	-	Ì	Ì	
114 115 116 117	E		G	6	+		-	+	T	+		+	-	A				+	4		+		+	Ť	+	+	÷			=	_		+		G			Ŀ				÷	C	Ì	+	-	Ŧ	Ŧ	
119	E		GI		-			c			ci	÷	ļ				G	Ť	-	T	+	+	+	+	1	Ŧ	T						Ŧ	-	+ G	+		C	+			- 1	0000			Ţ			
121	E		GI				- i	CI CI		1		+	T T						c	1	+	1	+	1		+	1			1			1					C				+	0000	-+				+	_
123 124 125 126		+ +	G		Ţ			ci	Ţ	+	C	+							C	+		+	-	+			-	-			1		1	-								-	CCCC	1		_			
126 127 128 129		11	G							-			+						C	+		+		+		+		-					+					С					C C C C C						_
130	E										C	+					q	T		+		+	+	+	+	1	-					1	+					C C	1				C C C C C C C C		+		+		
131 132 133 134 135 135	E		G	_	=					+	+	+	+	A			+	+	4	+	+	+	+	+	+	+	+	+	H	_	A	+		+			T	L	÷			Ì	0000		Ì		Ŧ	+	
136 137 138 139	E	i i	5				F			+	+	+	+	Ħ					-	İ			G	+		-	1	-	H				+					С	F			∃	000				G		_
140 141	-	1 1	G				A	c		+	+	+		T			G	+	C	+				-	+	T	1	-					-				1	CCC	÷				00000		C		T		_
142 143 144 145		H	- 4		Ħ		_M'_			1	1	Ì					+	1				1	Ŧ		+	+							Ì					C	+				CC	H			+	T	-
146	╋	- C	_ <u>(</u>							T	T		T	A		G			c		T	-				+	+				A	Ţ		+			I	0000	C				0000				+	+	_
149 149 150 151 152 153	F		G					C		T	C							+							-	İ	-	1							+ 6			i C	C C	1		+	C C	+			+		_
153			G							T	-	Ŧ	-		_		-	-		_	-	-	+	+	-		+	-		_						31	T		+			_	JOC		_	_		+	_

Fig 1

Table 4 HVI and HVII mtDNA polymorphism in Caucasian populations

Populations	Sample size	nb of different sequences found	nb of variable nucleotide sites	Total number of mutations	Mean pairwise difference	<i>P</i> *
SW Swiss (This study)	154	137	140	1094	8.58	0.0084
French [18]	50	49	88	345	8.25	0.0212
German [15, 16]	200	178	143	1274	7.90	0.0085
Austrian [17]	101	89	118	743	8.95	0.0131
British [13]	100	90	119	701	8.52	0.0129

\* Probability of two randomly selected individuals having identical mtDNA sequences

Table 5 Frequency (%) of the most common HVI and HVII mtDNA types in five Caucasian populations

mtDNA type	SW Swiss $n = 154$	French $n = 50$	German n = 200	Austrian $n = 101$	British $n = 100$	Total $n = 605$
263 G/315.1 C	2.59	4.0	3.5	3.96	4.0	3.88
263 G/309.1 C/315.1 C	1.29	0	5.0	1.98	2.0	2.91
152 C/263 G/315.1 C	0	0	1.5	0.99	3.0	1.16
16126 C/16294 T/16296 T/16304 C 73 G/263 G/315.1 C	1.95	0	1.0	0.99	0	1.16
16069 T/16126 C/16145 A/16231 C/16261 T 73 G/ 150 T/152 C/195 C/215 G/263 G/295 T 309.1 T/215.1 C/319 C	1.29	0	1.0	2.97	2.0	1.35

the electropherogram data by the Factura software using a 50% mixed base threshold: the base call was converted to the appropriate mixed base Y, R or M (IUB code) if the ratio of the lower peak to the highest peak reached the threshold value. These nucleotide mixtures occurred on both strands and were reproducible after repeated sequencing with the same method or with the dye primer method. These positions, described as point mutation heteroplasmy [6, 8, 9, 17, 28–35], were sometimes difficult to differentiate from various sequencing artefacts due to the frequently unbalanced nucleotide signals. But the use of the new BigDye Terminator sequencing chemistry [27], with improved sensitivity and peak signals allowed a more reliable detection of the heteroplasmic point mutation. This phenomenon is of great concern as recently a higher heteroplasmy rate has been reported in the control region of human DNA and is consistent with the rapid evolution and the high substitution rate of mtDNA [28–35].

In our Swiss database 11% of individuals were found to have a homopolymer C-stretch at position 16189 and the blurred sequencing pattern [36] was only present in 5.2% of the samples. Individuals with the 16189 variant and another variant in or near the homopolymer C-stretch (16186, 16187, 16192) did not show this phenomenon.

In HVII another C-stretch was present with insertions between position 309 (309.1, 309.2, 309.3) and 315 (315.1 and 315.2) [37]. In our population samples, 61.66% of 310 C-stretch length polymorphism were observed. In rare cases, in which the number of C was higher than 8, it was impossible to give the correct number and the following sequence became unreadable. For the two regions HVI and HVII, the sequences obtained for these length heteroplasmic samples were always confirmed by a second independent sequencing of the same strand.

## Population data

In our population sample of 154 random individuals, 137 HVI and HVII mtDNA types were observed (Fig. 1). Of these, 124 sequences were observed in only one individual whereas 10 sequences were observed in 2 individuals, 2 sequences in 3 individuals and 1 sequence in 4 individuals. Compared to the Anderson sequence [2], the majority of the variable sites in HVI and HVII were transitions whereas insertions were particularly frequent in the HVII region. The mean pairwise difference was 8.58. The probability of two unrelated individuals having the same sequence was 0.84%.

The genetic data from our population was similar to those calculated for four other Caucasian populations [13, 15–18] (Table 4). The mean pairwise difference calculated for the HVI region of our South West Swiss population was 4.69 whereas Pult et al. [14] gave a mean pairwise difference of 3.63 for a sample of 74 native Swiss individuals. The difference between these two mean pairwise values can be explain by two factors. First, the Pult et al. data included 15 related german speaking inhabitants of a small Alp village situated in the Italian area of Switzerland (W. Schaffner, personnal communication). Secondly, our South West Swiss population samples were taken from three different french speaking cantons corresponding to the area of Switzerland served by our Institute. Considering the distribution of the five most common mtDNA sequences in these four Caucasian populations [13, 15–18] and our population (Table 5), we observe that two sequences (263G / 315.1 C and 263G / 309.1 C / 315.1 C) showed a frequency between 1.29 and 4% in all five populations whereas two others were missing each time in two populations and the last sequence was only missing in the French population (n = 50). As previously described [17, 38], these results demonstrate that the control region sequence HVI and HVII is long enough to discriminate even within closely related mtDNA in a population, thus concealing enhanced regional matching probability in the sample sizes considered here.

## Casework samples

Samples from forensic cases (113 samples from 18 cases) were extracted and typed: 69 evidence samples and 44 blood or saliva reference samples. Despite the fact that 55% of the evidence samples yielded a negative result for nuclear DNA with the human dot quantitation system, the mtDNA results of the nested PCR were positive in 81.2% of the samples (Table 2). Considering the cycle sequencing analysis, conclusive results were obtained for 87.5% of the samples. The mtDNA success rates of different biological evidence samples are reported in Table 6. The inconclusive results of the sequencing analysis were mainly due to the presence of several ambiguities or mixed pattern. As shown in Table 6, hairs were the most evidence samples. All the hairs tested were shed hairs and for each forensic case, at least two different hairs or two segments of the same hair were analysed. We always obtained the same results for the duplicate testing and we never observed heteroplasmic point mutation. Whenever possible, it is necessary to test several hairs because hairs may show different levels of heteroplasmy and sometimes different mtDNA types within the same individual [33, 34]. This situation is representative of the developmental origins of hair follicles as hairs formation originate from discrete groups of stem cells [33]. The reference samples were excluded in 63.65% of the cases. All exclusions were based on two or more base differences with no evidence of point or length heteroplasmy. That is relevant in forensic identification of remains or biological stains as mtDNA is often use to compare questioned samples to presumed maternal reference and the high substitution rate can induce differences between true maternal relatives [29, 35].

Concerning the included reference samples, we compared the observed sequences with the 5 Caucasian populations mentioned in this work (Table 4) and we reported as frequency in our evidence cases the number of the observed sequence in these 605 individuals. Two sequences had a frequency of 10/605 and 2 others 6 /605 whereas 12 sequences were never observed in these 605 Caucasian samples.

In conclusion, the mtDNA sequencing technique using nested PCR and fluorescence capillary electrophoresis was very useful to resolve the identification of difficult forensic casework knowing the different problems related to the direct sequencing method and the mtDNA genetic.

**Acknowledgements** The authors wish to thank Mrs C. Jan and Mr C. Gregoretti for their technical assistance. They would also thank Dr. G. van Melle for computing the mean pairwise differences.

 Table 6
 mtDNA sequencing success rate of different biological evidence samples

Sample type	Sample number	Success rate (%)
Hair shaft	22	90.9
Sweat stain	16	62.5
Blood stain	15	80.0
Saliva stain	6	83.3
Bone	4	100.0
Nail	4	100.0
Serum	1	100.0
Muscle	1	100.0

## References

- Bogenhagen D, Clayton D (1974) The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. J Biol Chem 249:7991–7995
- Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465
- Giles RE, Blanc H, Cann HM, Wallace DC (1980) Maternal inheritance of human mitochondrial DNA. Proc Natl Acad Sci USA 77:6715–6719
- Sullivan KM, Hopgood R, Gill P (1992) Identification of human remains by amplification and automated sequencing of mitochondrial DNA. Int J Legal Med 105:83–86
- Holland MM, Fisher DL, Mitchell LG, Rodriquez WC, Canik JJ, Merril CR, Weedn VW (1993) Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam War. J Forensic Sci 38:542–553
- 6. Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N, Tully G, Evett I, Hagelberg E, Sullivan K (1994) Identification of the remains of the Romanov family by DNA analysis. Nat Genet 6: 130–135
- Holland MM, Fisher DL, Roby RK, Ruderman J, Bryson C, Weedn VW (1995) Mitochondrial DNA sequence analysis of human remains. Crime Lab Digest 20:109–115
- Ivanov PL, Wadhams MJ, Roby RK, Holland MM, Weedn VW, Parsons TJ (1996) Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. Nat Genet 12: 417–420
- Lutz S, Weisser HJ, Heizmann J, Pollak S (1996) mtDNA as a tool for identification of human remains. Int J Legal Med 109: 205–209
- 10. Allen M, Engström A, Meyers S, Handt O, Saldeen T, von Haeseler A, Pääbo S, Gyllensten U (1998) Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: sensitivity and matching probabilities. J Forensic Sci 43:453–464
- Weichhold GM, Bark JE, Korte W, Eisenmenger W, Sullivan KM (1998) DNA analysis in the case of Kaspar Hauser. Int J Legal Med 111:287–291
- 12. Pfeiffer H, Steighner R, Fisher R, Mörnstad H, Yoon CL, Holland MM (1998) Mitochondrial DNA extraction and typing from isolated dentin-experimental evaluation in a Korean population. Int J Legal Med 111:309–313
- Piercy R, Sullivan KM, Benson N, Gill P (1993) The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. Int J Legal Med 106:85–90

- 14. Pult I, Sajantila A, Simanainen J, Georgiev O, Schaffner W, Pääbo S (1994) Mitochondrial DNA sequences from Switzerland reveal striking homogeneity of European populations. Biol Chem Hoppe Seyler 375:837–840
- 15. Lutz S, Weisser HJ, Heizmann J, Pollak S (1998) Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. Int J Legal Med 111:67–77
- 16. Lutz S, Weisser HJ, Heizmann J, Pollak S (1998) Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany:Errata. Int J Legal Med 112: 145–150
- 17. Parson W, Parsons TJ, Scheithauer R, Holland MM (1998) Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case. Int J Legal Med 111:124–132
- Rousselet F, Mangin P (1998) Mitochondrial DNA polymorphisms: a study of 50 French Caucasian individuals and application to forensic casework. Int J Legal Med 111:292–298
- 19. Lee SD, Shin CH, Kim KB, Lee YS, Lee JB (1997) Sequence variation of mitochondrial DNA control region in Koreans. Forensic Sci Int 87:99–116
- 20. Walsh PS, Metzger DA, Higuchi R (1991) Chelex 100 as medium for simple extraction of DNA for PCR-based typing from forensic material. Bio techniques 10:506–513
- Sullivan M, Hopgood R, Lang B, Gill P (1991) Automated amplification and sequencing of human mitochondrial DNA. Electrophoresis 12:17–21
- 22. Venigalla BR (1994) Direct sequencing of polymerase chain reaction-amplified DNA. Anal Biochem 216:1–14
- 23. Rieder MJ, Scott LT, Tobe VO, Nickerson DA (1998) Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mito-chondrial genome. Nucleic Acids Res 26:967–973
- 24. Leonard JT, Grace MB, Buzard GS, Mullen MJ, Barbagallo CB (1998) Preparation of PCR products for DNA sequencing. Biotechniques 24:314–317
- 25. Wilson MR, Stoneking M, Holland MM, DiZinno JA, Budowle B (1993) Guidelines for the use of mitochondrial DNA sequencing in forensic science. Crime Lab Digest 20:68–77
- 26. Wilson MR, DiZinno JA, Polanskey D, Replogle J, Budowle B (1995) Validation of mitochondrial DNA sequencing for forensic casework analysis. Int J Legal Med 108:68–74

- 27. Rosenblum CC, Lee LG, Spurgeon SL, Kahn SH, Menchen SM, Heiner CR, Chen SM (1997) New dye-labeled terminators for improved DNA sequencing patterns. Nucleic Acids Res 25:4500– 4504
- 28. Comas D, Pääbo S, Bertrandpetit J (1995) Heteroplasmy in the control region of human mitochondrial DNA. Genome Res 5: 89–90
- 29. Jazin EE, Cavelier L, Eriksson I, Oreland L, Gyllensten U (1996) Human brain contains high levels of heteroplasmy in the non coding regions of mitochondrial DNA. Proc Natl Acad Sci USA 93:1282–1287
- Bendall KE, Macaulay VA, Baker JR, Sykes BC (1996) Heteroplasmic point mutations in the human mtDNA control region. Am J Hum Genet 59:1276–1287
- 31. Parsons TJ, Muniec DS, Sullivan K, Woodyatt N, Alliston-Greiner R, Wilson MR, Berry DL, Holland KA, Weedn VW, Gill P, Holland MM (1997) A high observed substitution rate in the human mitochondrial DNA control region. Nat Genet 15: 363–368
- 32. Lightowlers RN, Chinnery PF, Turnbull DM, Howell N (1997) Mammalian mitochondrial genetics: heredity, heteroplasmy and disease. Trends in Genet 13:450–455
- 33. Bendall KE, Macaulay VA, Sykes BC (1997) Variable levels of a heteroplasmic point mutation in individual hair roots. Am J Hum Genet 61:1303–1308
- 34. Wilson MR, Polanskey D, Replogle J, DiZinno JA, Budowle B (1997) A family exhibiting heteroplasmy in the human mitochondrial DNA control region reveals both somatic mosaicism and pronounced segregation mitotypes. Hum Genet 100:167– 171
- 35. Hühne J, Pfeiffer H, Brinkmann B (1998) Heteroplasmic substitutions in the mitochondrial DNA control region in mother and child samples. Int J Legal Med 112:27–30
- 36. Bendall KE, Sykes BC (1995) Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. Am J Hum Genet 57:248–256
- 37. Hauswirth WW, Clayton DA (1985) Length heterogeneity of a conserved displacement-loop sequence in human mitochondrial DNA. Nucleic Acids Res 22:8093–8104
- Melton T, Wilson M, Batzer M, Stoneking M (1997) Extent of heterogeneity in mitochondrial DNA of European populations. J Forensic Sci 42:437–446