Int J Legal Med (2001) 114 : 224–228

ORIGINAL ARTICLE

Adriano Tagliabracci · Chiara Turchi
Loredana Buscemi · Corrado Sassaroli

Polymorphism of the mitochondrial DNA control region in Italians

Received: 14 February 2000 / Accepted: 23 May 2000

Abstract Mitochondrial DNA sequences of the hypervariable regions HVI and HVII were analysed in 83 Caucasians living in central Italy to expand the database for forensic identification purposes, and 75 different haplotypes resulting from 62 polymorphic positions in HVI and 44 in HVII were observed. The most frequent haplotype (263G, 309.1C, 315.1C) was shared by 7 individuals, 2 haplotypes were shared by 2 individuals, and 72 were unique. The genetic diversity was found to be 0.99 and the random match probability 1.9%. A condition of sequence heteroplasmy was found in only one case at nt 16311, whereas a length heteroplasmy was found in the homopolymeric stretch of cytosines 303–315. Our results indicate that in direct sequencing beyond the poly-cytosine stretch, the overlap is due to length heteroplasmy, whereas the blurred signal occurs when the stretch is composed of more than 10 cytosines.

Keywords Mitochondrial DNA · PCR · Sequencing · Population frequencies · Heteroplasmy · Forensic identification

Introduction

Mitochondrial DNA (mtDNA) is a circular molecule of 16569 bp, present in up to thousands of copies per cell. The high polymorphism of the D-loop (control region) in the three hypervariable segments HV1, HV2 and HV3 and the number of copies for each cell make mtDNA a first-choice marker in identification analysis performed in cases where DNA is only present in small amounts or where it is badly degraded (Pääbo et al. 1988; Ginther et al. 1992; Holland et al. 1993; Gill et al. 1994; Krings et al. 1997; Weichhold et al. 1998). Maternal inheritance and absence of recombination are other specific features of this molecule, useful when identifying human remains by comparison with matrilineal descendants or in establishing the maternal relationship of living individuals. Since the early 1990s the two non-coding HV1 and HV2 fragments of the control region have been analysed in an increasing number of types of biological evidence from human remains, and mtDNA sequencing strategies have been validated for forensic applications (Ginther et al. 1992; Hopgood et al. 1992; Sullivan et al.1992; Fisher et al. 1993; Holland et al. 1993; Gill et al. 1994; Wilson et al. 1995; Hopwood et al. 1996). Databases of mtDNA sequences were also established from various populations to assess frequencies of haplotypes and to estimate random match probabilities. However, the absence of recombination in mtDNA that could affect the probability of identity by chance (Allen et al. 1998), the occurrence of population-specific lineage groups (Torroni et al. 1998), and the observation that the number of haplotypes increases with sample size (Budowle et al. 1999) suggest the need to expand databases as much as possible. Unlike the theoretical model proposed by Francalacci et al. (1996), in which haplotypes approached the level of saturation, the increase in number of alleles did show a nearly linear trend in a large sample sequenced by Pfeiffer et al. (1999). The main aim of the present work was to contribute to the expansion of HVI and HVII sequence mtDNA databases by investigating a sample of Caucasians from central Italy.

Materials and methods

Extraction

DNA from fresh whole blood of 83 unrelated subjects of Caucasian origin born in central Italy was extracted according to standard procedures by phenol-chloroform (Budowle and Baechtel

Supplementary material Data on the 75 mtDNA haplotypes found in the 83 Italians in this population study (Table S1) are available in electronic form on Springer-Verlag's server at http://link.springer.de/link/service/journals/00414/index.htm

A. Tagliabracci (✉) · C. Turchi · L. Buscemi · C. Sassaroli
Istituto di Medicina Legale, Università di Ancona,
Policlinico Torrette, 60020 Ancona, Italy
e-mail: tagliabr@popcsi.unian.it;
Fax: +39-071-5964723

Table 1 Sequences of primers used in this study

Regions	Amplification primer	Nucleotide sequence
HV1	L15971	5′-TTAACTCCACCATTAGCACC-3′
	H16414	5′-CACGGAGGATGGTGGTCAAG-3′
	L16140	5′-TACTTGACCACCTGTAGTAC-3′
	H16258	5′-TGGCTTTGGAGTTGCAGTTG-3′
HV2	L00015	5′-CACCCTATTAACCACTCACG-3′
	H00389	5′-CTGGTTAGGCTGGTGTTAGG-3′
	L00145	5′-CTCATCCTATTATTTATCGC-3′
	H00274	5′-TGTGTGGAAAGTGGCTGTGC-3′
	Sequencing primer	Nucleotide sequence
HV1	F15971-Seq	5′-TTAACTCCACCATTAGCACC-3′
	R16414-Seq	5′-CACGGAGGATGGTGGTCAAG-3′
HV2	F29-Seq	5′-CTCACGGGAGCTCTCCATGC-3′
	R381-Seq	5′-GCTGGTGTTAGGGTTCTTTG-3′

1990), followed by slot blot quantitation using a specific higher primate probe (Waye et al. 1989).

Amplification

The two hypervariable regions of mtDNA were amplified by the polymerase chain reaction (PCR) using the primers listed in Table 1.

PCR was carried out in a 50 µl reaction mixture with 1 × reaction buffer (10 mM Tris-HCl pH 8.3, 50 mM KCl, 0.1% Triton X-100), 1.5 mM $MgCl_2$, 200 mM each dNTP, 0.4 mM each primer, 2.5 U Taq polymerase and 0.1–1 ng DNA template was added to the reaction mixture. Amplification was carried out in a GenAmp System 2400 thermal cycler (Perkin Elmer, Norwalk, Conn.) using a hot start at 94 °C for 30 s, followed by 38 cycles at 94 °C for 20 s, 56 °C for 10 s, and 72 °C for 30 s. After purification with a Microcon 100 spin-dialysis columns (Amicon Inc. Beverly, Mass.), PCR products were quantified in a 2% agarose gel.

Sequencing and electrophoresis

Cycle sequencing was performed on both strands in a GenAmp System 2400 thermal cycler using the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems, Foster City, Calif.). An aliquot of 30 ng DNA and 2.5 pM sequencing primer (Table 1) were added to 4 µl reaction mix taken to 20 µl with deionized water. Sequencing was carried out at 96 °C for 15 s, 50 °C for 5 s, and 60 °C for 60 s, for 25 cycles. The removal of excess dye-deoxy terminators, primers and buffer was accomplished with Centri-Sep spin columns (Princeton Separations, Adelphia, N.J.). The vacuum-dried sequencing product was then resuspended in 12 µl of Template Suppression Reagent (Applied Biosystems), heat-denatured and chilled, and analysed by capillary electrophoresis (CE) on an ABI Prism 310 (Applied Biosystems). The capillary was filled with run buffer Performance Optimized Polymer POP 6 (PE/ABD). Samples were electrokinetically injected in 30 s at 2.0 kv and separated for 36 min at 416 V cm^{-1} and 50 °C. The resulting data were analysed with PE/ABD software Sequencing Analysis 3.0, and sequences were aligned and compared with the Cambridge sequence (Anderson et al. 1981) from nt 15991 to 16394 for HVI and from nt 49 to 361 for HVII by the Sequence Navigator Software 1.0.1 (Perkin Elmer).

Results and discussion

Primers for the HV1 and HV2 regions used in this study yielded fragments of 404 bp and 313 bp, respectively. Sequences from both fragments were automatically analysed, and a base call was assigned when peaks were clearly defined and could not be confused with minor peaks from background noise.

Seventy-five haplotypes resulting from 62 polymorphic positions in HV1 and 44 in HV2 sequences were observed in the 83 Italians in this study. The most frequent haplotype was shared by 7 individuals, 2 other less frequent haplotypes were shared by only 2 individuals each, and 72 were unique (Table S1). Comparison with the Cambridge reference sequence (Anderson et al. 1981) showed one A to G transition at nucleotide 263 in all subjects, and a stretch of more than 5 cytosines at 311–315 nucleotides in 79 individuals. Transitions were predominant in determining sequence deviation with respect to transversions and insertions. More pyrimidine than purine transitions were observed in HV1, a phenomenon explained by the chemical nature of the nucleotide and its position within the region (Wakeley 1993). The HV2 sequence only was involved in C insertions.

The haplotype 263G, 309.1C and 315.1C was the most frequent in this study and, when compared with other populations, it also turned out to be the most frequent in Caucasoids living in north-west Germany (Pfeiffer et al. 1999) and the second most frequent haplotype in US Caucasians (Budowle et al. 1999). Haplotype 263G, with one or more insertions at nt 309.1C and/or 315.1C, has also been found in other European populations (Piercy et al. 1993; Francalacci et al. 1996; Lutz at al. 1998; Parson et al. 1998; Rousselet and Mangin 1998) and may be considered a common trait in Caucasians.

The mean pair-wise difference found in this study (8.4) is very similar to that observed for other Caucasian populations, i.e. 8.46 for Tuscany (Francalacci et al. 1996), 8.48 for Britain (Piercy et al. 1993), 8.38 for France (Rousselet and Mangin 1998), 7.2 for France, 7.9 for America and 8.4 for Austria (Budowle et al. 1999). The genetic diversity calculated according to Tajima (1989) was 0.99, and the random match probability calculated according to Stoneking et al. (1991) was 1.9%. In view of the different sample sizes and the relative number of alleles, these values rank with Caucasian populations as reported by Budowle et al. (1999).

From the analytical point of view, the forward primer produces better sequencing reactions, as previously reported (Parson et al. 1998), depending on the purine/pyrimidine content in the two strands.

The major problem encountered in analysis was interpretation of sequence reactions in samples with overlapping peaks of approximately the same height or with unbalanced peaks still clearly above background noise, simulating a condition of heteroplasmy. This phenomenon, in which two or more mtDNA populations are present in a single individual, has shown a significant incidence in

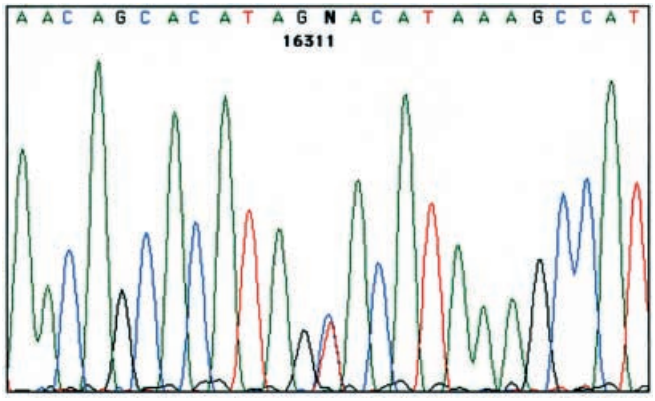

Fig. 1 Overlap of two peaks at nucleotide 16311, indicating a condition of heteroplasmy. Electropherogram displays one N, due to a mixture of T and C nucleotides at approximately the same ratio

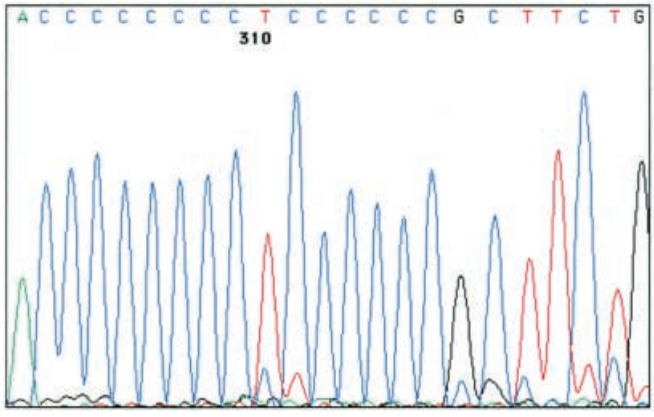

Fig. 2 Two sequences in the C-stretch beyond nt 310, due to length heteroplasmy. Electropherogram shows another minor sequence over baseline, starting at nt 310, at a very low level and shifted by one base with respect to the major sequence

several studies (Bendall et al. 1997; Parsons et al. 1997; Parson et al. 1998; Baasner et al. 1998; Hühne et al. 1998; Seo et al. 1998). Following the criteria suggested to differentiate artefacts from true substitutions in direct sequencing, i.e. replication of sequences (Parsons et al. 1997), sequence analysis of the opposite strand (Parson et al. 1998) and proportions of the two peaks (Hühne et al. 1998), all ambiguous sequences were correctly defined and only one piece of dental pulp evidence from casework showed a sequence exhibiting a heteroplasmic variant at nt 16311 (Fig. 1).

Heteroplasmic mutation points have been described, especially in the homopolymeric tracts of cytosines between nt 16184–16193 and 303–315, where a T to C transition, at nt 16189 in HV1 and nt 310 in HV2, determines sequence heteroplasmy with a relatively common frequency (Bendall and Sykes 1995; Marchington et al. 1997). Probably due to a replication slippage mechanism, this mutation is almost always associated with one or more cytosine insertions in the polycytosine stretches which produce length heteroplasmy, and in direct sequencing beyond these tracts a characteristic unclear sequence is usually found (Bendall and Sykes 1995). The unclear sequence beyond the C-stretch is believed to be determined by sequencing artefacts or by overlap of different sequences deriving from length heteroplasmy (Bendall and Sykes 1995).

The T to C transition at nt 16189 was observed in 16 subjects (19%) out of the 83 samples of this study, in accordance with the literature (Bendall and Sykes 1995; Hühne et al. 1998; Parson et al. 1998) and due to the simultaneous C to T transition in nt 16186 or C to A transversion in nt 16184 or 16192 affecting some sequences, a stretch of 10 or more cytosines was found in only 10 subjects. Of these, the transition at nt 16189 was coupled with an A to C transversion at nt 16183 in 5 sequences, and there was a double A to C transversion at nt 16182 and 16183 in one sequence (Table S1), generating 11 and 12 cytosine stretches, respectively; in the remaining 4 samples the stretch was 10 cytosines long. In the cases where 11 or 12 cytosine stretches were present, the signal beyond the homopolymeric C-stretch dropped drastically. In homopolymeric C-tracts 303–309 and 311–315, almost all samples showed one or more C insertions, but the T at nt 310 was constantly present and the stretch was 9 cytosines long at most. The signal of the sequences following this stretch was clear and two overlapping sequences could be interpreted in 14 samples (Fig. 2), with generally small minor peaks, as also reported by Parsons et al. (1997). Moreover, one mother-son pair from casework (data not shown) showed an unusual T to C transition at position 310, generating a stretch of 13 nucleotides, and in this case beyond the C-stretch the sequence was unclear, as in HV1 C-stretch.

From the results of our study, it is assumed that the blurred signal in direct sequencing of both regions is related exclusively to the length of the C-stretch, as the critical length for signal drop-out is more than 10 cytosines. Direct sequencing does not establish whether the blurred signal is due to a sequencing artefact or length heteroplasmy, but Bendall and Sykes (1995) ascertained by cloning that this phenomenon is due to length heteroplasmy. The length heteroplasmy derived from cytosine insertions in the homopolymeric C-tract 303–315 observed in our sample in direct sequencing may be considered as two overlapping but clearly distinguishable sequences (Fig. 2).

An A to C transversion in the four adenines following the C-stretch and a C to T transition at nt 16186, 16187 or 16192 were described by Bendall and Sykes (1995) and interpreted as factors limiting homopolymeric tract expansion and length heteroplasmy since, according to these authors, it is improbable that replication slippage acts on homopolymeric tracts containing eight or fewer consecutive cytosines. Our study showed the presence of mutations in the adjacent region and inside the stretch, and confirmed these overall dynamics in the homopolymeric tracts of cytosines.

Another problem encountered in analysis, affecting only 12 samples from our study, was the overlapping of

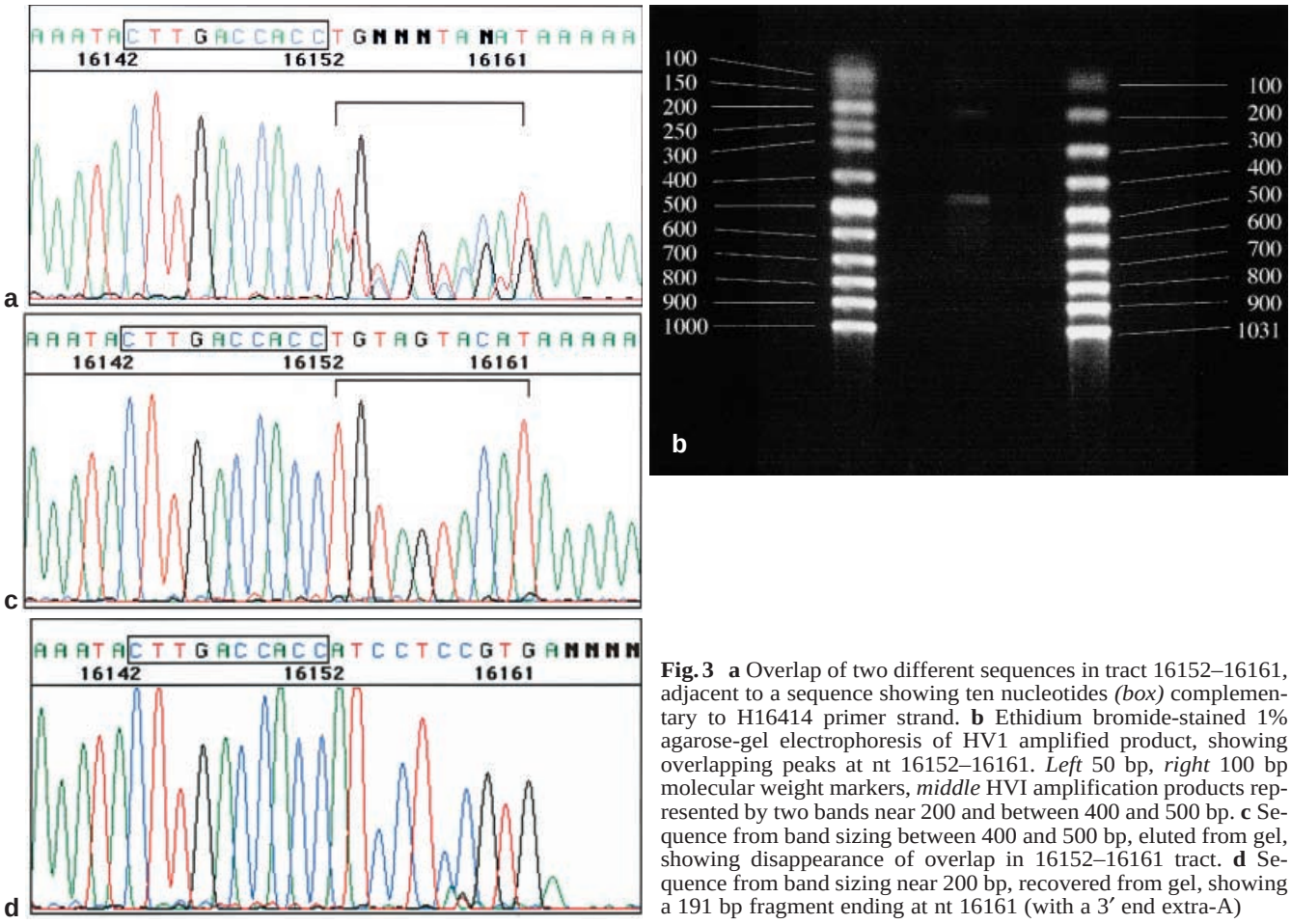

Fig. 3 **a** Overlap of two different sequences in tract 16152–16161, adjacent to a sequence showing ten nucleotides *(box)* complementary to H16414 primer strand. **b** Ethidium bromide-stained 1% agarose-gel electrophoresis of HV1 amplified product, showing overlapping peaks at nt 16152–16161. *Left* 50 bp, *right* 100 bp molecular weight markers, *middle* HVI amplification products represented by two bands near 200 and between 400 and 500 bp. **c** Sequence from band sizing between 400 and 500 bp, eluted from gel, showing disappearance of overlap in 16152–16161 tract. **d** Sequence from band sizing near 200 bp, recovered from gel, showing a 191 bp fragment ending at nt 16161 (with a 3′ end extra-A)

two sequences for 10 nucleotides in the 16152–16161 stretch of the HVI region (Fig. 3a), where a minimal amount of DNA was recovered from old bloodstains and hairs (data not shown). Since the adjacent region 16142–16151 shows a sequence identical to those where the first 10 nucleotides of the reverse H16414 primer anneal, and because from this overlap the peaks showed a drastic decrease in height up to the end of the molecule, we believe that incomplete reverse primer annealing occurred in this region, which yielded an additional 191 bp long PCR product. Agarose gel electrophoresis of the HVI amplification products from these samples confirmed the presence of two bands, differing for the postulated length (Fig. 3b). Two well-defined sequences without any overlap in the critical region 16152–16161 (Fig. 3c, d), obtained by sequencing the recovered bands from agarose gels, further validated the hypothesis, as well as the disappearance of the overlap when a new set of primers was used (L15997 and H16401), with only four bases of the reverse primer complementary to the homology region at nt 16148–16151. High stringency amplification of this region and attempts to vary the concentrations of primers and DNA template were not able to resolve the problem of overlapping, although the hypothesis of the unbalanced concentration of the amplification reaction was the more probable explanation, as only old or degraded samples were affected.

In conclusion, a sample of 83 subjects from central Italy was examined in order to expand the mtDNA database for identification purposes. These data confirm the high polymorphism of mtDNA observed in studies performed until now. Analysis of casework also confirmed the suitability of this molecular analysis in identification when nuclear DNA fails.

Acknowledgements This work was supported by a grant from MURST (9906404127–002).

References

Allen M, Engstrom A-S, Meyers S, Handt O, Saldeen T, von Haeseler A, Paabo 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

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

Baasner A, Schäfer, Junge A, Madea B (1998) Polymorphic sites in human mitochondrial DNA control region sequences: population data and maternal inheritance. Forensic Sci Int 98: 169–178

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

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

Budowle B, Baechtel FS (1990) Modifications to improve the effectiveness of restriction fragment length polymorphism typing. Appl Theor Electrophor 1: 181–187

Budowle B, Wilson MR, DiZinno JA, Stauffer C, Fasano MA, Holland MM, Monson KL (1999) Mitochondrial DNA regions HVI and HVII population data. Forensic Sci Int 103: 23–35

Fisher DL, Holland MM, Mitchell L, Sledzik PS, Wilcox AW, Wadhams M, Weedn VW (1993) Extraction, evaluation, and amplification of DNA from decalcified and un-decalcified United States Civil War bones. J Forensic Sci 38: 60–68

Francalacci P, Bertranpetit J, Calafell F, Underhill P (1996) Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. Am J Phys Anthropol 100: 443–460

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

Ginther C, Issel-Tarver L, King MC (1992) Identifying individuals by sequencing mitochondrial DNA from teeth. Nat Genet 2: 135–138

Holland MM, Fisher DL, Mitchell LG, Rodriguez 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

Hopgood R, Sullivan KM, Gill P (1992) Strategies for automated sequencing of human mitochondrial DNA directly from PCR products. Biotechniques 13: 82–92

Hopwood AJ, Mannucci A, Sullivan KM (1996) DNA typing from human faeces. Int J Legal Med 108: 237–243

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

Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Paabo S (1997) Neandertal DNA sequences and the origin of modern humans. Cell 90: 19–30

Lutz S, Weisser H-J, 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

Marchington DR, Hartshorne GM, Barlow D, Poulton J (1997) Homopolymeric tract heteroplasmy in mtDNA from tissues and single oocytes: support for a genetic bottleneck. Am J Hum Genet 60: 408–416

Pääbo S, Gifford JA, Wilson AC (1988) Mitochondrial DNA sequences from a 7000-year-old brain. Nucleic Acids Res 16: 9775–9778

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

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

Pfeiffer H, Brinkmann B, Hühne, Rolf B, Morris AA, Steighner R, Holland MM, Forster P (1999) Expanding the forensic German mitochondrial DNA control region database: genetic diversity as a function of sample size and microgeography. Int J Legal Med 112: 291–298

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

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

Seo Y, Stradmann-Bellinghausen B, Rittner C, Takahama K, Schneider PM (1998) Sequence polymorphism of mitochondrial DNA control region in Japanese. Forensic Sci Int 97: 155–164

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

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

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585–595

Torroni A, Bandelt H-J, D' Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, Forster P, Savontaus M-L, Bonné-Tamir B, Scozzari R (1998) mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. Am J Hum Genet 62: 1137–1152

Wakeley J (1993) Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. J Mol Evol 37: 613–623

Waye JS, Presley L, Budowle B, Shutler GG, Fourney RM (1989) A simple method for quantifying human genomic DNA in forensic specimen extracts. Biotechniques 7: 852–855

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

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