Int J Legal Med (2004) 118 : 47–54
DOI 10.1007/s00414-003-0418-z

ORIGINAL ARTICLE

**Anita Brandstätter · Harald Niederstätter**
**Walther Parson**

# Monitoring the inheritance of heteroplasmy by computer-assisted detection of mixed basecalls in the entire human mitochondrial DNA control region

Received: 21 August 2003 / Accepted: 27 November 2003 / Published online: 9 January 2004


**Abstract** The entire mitochondrial DNA control region (~1122 bp) of 270 blood samples (135 mother-child pairs) was determined by direct sequencing. Overall, 135 'generational events' were screened and within these, 20 mother-offspring pairs demonstrated more than 1 mtDNA haplotype. In 13 families, differences in the haplotypes between mother and offspring were detected in the form of heteroplasmic substitutions. Intergenerational comparisons led to the identification of three heteroplasmic point mutations and eight heteroplasmic length mutations affecting the children only. In two cases, a point heteroplasmy of the maternal sequence was resolved to homoplasmy in the corresponding sequence of the child. These discordant maternal-offspring haplotypes suggest that the shift in the mtDNA haplotype was the result of segregation of a limited maternal subpopulation of mtDNA. As technical implement, quality values assigned to basecalls were tested for their application in automated point heteroplasmy detection.

**Keywords** Point heteroplasmy · Length heteroplasmy · Genetic bottleneck · Sequence quality values · Sequence analysis · Forensic science

## Introduction

Studies of human mitochondrial DNA (mtDNA) transmission in families have revealed that multiple mtDNA haplotypes (heteroplasmy) may be present in the same individual [1, 2, 3]. The level of heteroplasmy can vary among family members and among different tissues within the same individual [4]. The consensus view is that human mitochondrial DNA is exclusively inherited down the maternal line [5]. As a result, all the offspring of a single female are clonal for mtDNA. Whether partitioning of the existing mtDNA molecules occurs by random drift or through a sorting mechanism that determines the distribution of organelles to daughter cells is still unclear.

The lack of recombination is fundamental in the forensic application of mtDNA and is critical in understanding the transmission of diseases due to mtDNA mutations. There is an ongoing discussion between groups who favour the recombinational explanation of mtDNA variation in humans [6, 7, 8, 9, 10] and others, who question this way of interpretation [11, 12, 13, 14, 15]. Despite the high mtDNA copy number (approximately 100,000 or more) in mature oocytes, and despite the relatively small number of cell divisions during oogenesis, mtDNA sequence variants segregate rapidly between generations [16, 17]. This astonishing behaviour has been ascribed to the presence of an mtDNA 'bottleneck' in oogenesis or early embryogenesis. Selective and/or replicative advantages of some mutations combined with a severe bottleneck during the mitochondrial segregation accompanying mitosis, are the mechanisms probably involved in the origin of mitochondrial genome instability [18, 19].

Compared to genomic DNA, the mitochondrial DNA control region shows an especially high mutation rate [20]. The high mutation rate and clonal maternal transmission lead to the accumulation of novel mtDNA mutations within the population [2]. Knowledge of the frequencies of new mutations in maternal lineages is of central importance for the application of mtDNA in forensic identity investigations. In recent years, there have been reports on the transmission of mtDNA heteroplasmy located in the two hypervariable regions (HV1 and HV2) of the mtDNA control region [21, 22]. The present study attempts to add to the understanding of the assumed genetic bottleneck by completing existing information with the analysis of the entire mitochondrial control region (~1122 bp).

In addition, we examined the utility of quality values assigned to basecalls in order to distinguish mixed basecalls (e.g. heteroplasmy) from clear basecalls.

A. Brandstätter · H. Niederstätter · W. Parson (✉)
Institute of Legal Medicine, University of Innsbruck,
Müllerstrasse 44, 6020 Innsbruck, Austria
Tel.: +43-512-507-3303, Fax: +43-512-507-2764,
e-mail: Walther.Parson@uibk.ac.at

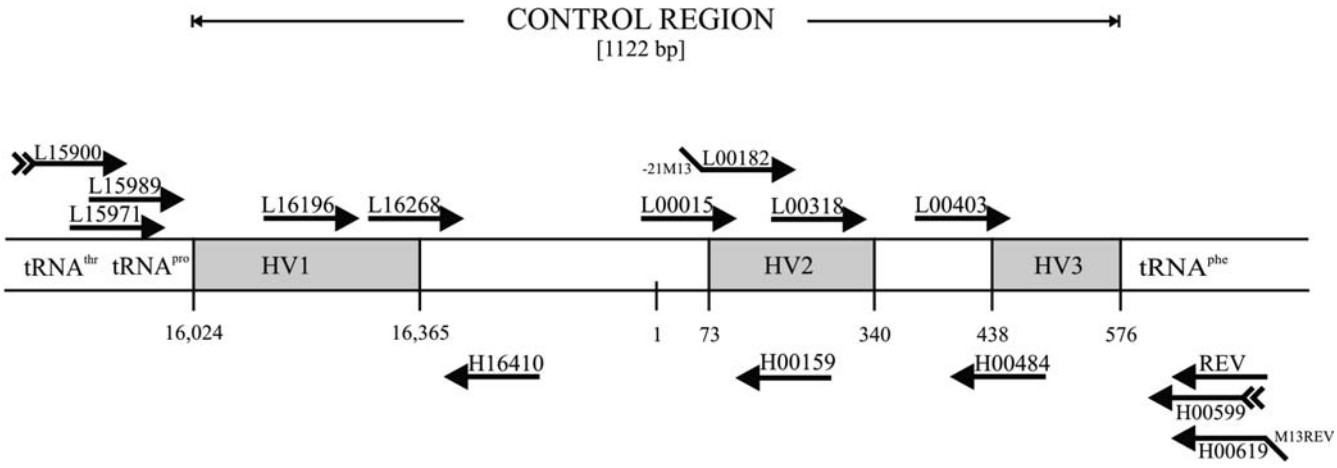


**Fig. 1** Amplification and sequencing strategies for the analysis of the human mtDNA control region. Primers used in the amplification of the CR and in the separate amplification of HV3 are shown as *shafted arrows*. The location of the primers is indicated by the position in the rCRS that corresponds to the 5'-end of the primer except for L16169 and L00182, which are numbered according to the location of their 3'-ends in the rCRS. "–21M13" and "M13REV" designate the sequences derived from the universal forward or reverse primer which were added to the 5'-terminus of the PCR primers

## Materials and methods

DNA samples and extraction

DNA was extracted from blood obtained from 135 unrelated Austrian Caucasian women and their children (one child per woman) using Chelex 100 [23].

Amplification and sequencing of mtDNA

PCR amplification using the primer pair L15900/H599 yielded a fragment of approximately 1250 bp, which covers the entire mitochondrial control region. The scheme for the amplification strategy is depicted in Fig. 1, all primer sequences are given in Table 1. PCR was performed in a total volume of 25 µl consisting of 1× PCR reaction buffer (Clontech, BD Biosciences, Palo Alto and Mountain View, CA), 200 µM each dNTP, 0.5 µM each primer and 1× Advantage II polymerase (Clontech). Amplification was carried out on a 9700 GeneAmp Thermal Cycler (Perkin Elmer-Applied Biosystems, AB, Foster City, CA). The reaction cocktails were heated to 95°C (2 min) and then put through 35 reaction cycles: 95°C for 15 s, 56°C for 30 s and 72°C for 90 s, followed by a final extension phase at 72°C for 10 min. Amplicons were purified by ultrafiltration (Multiscreen, Millipore, Billerica, MA) and directly sequenced by cycle sequencing with BigDye Terminator sequencing reagents (AB). Reaction was carried out in a final volume of 10 µl containing 2 µl BigDye Terminator RR mix, 2 µl halfBD Dye Terminator Sequencing Reagent (Sigma-Aldrich, St. Louis, MO), 1.6 pmol primer and 4 µl PCR product. Cycling was performed (after a first denaturation step of 95°C, 2 min) for 30 cycles of 15 s at 95°C, 10 s at 50°C, and 4 min at 60°C. Each template was sequenced in the forward direction with primers L15971, L16268, L00015, L00403, and in the reverse direction with H16410,

**Table 1** Sequences of primers used for amplification and sequencing of mitochondrial DNA segments in this study. "L" and "H" stand for the light and heavy strand of the mtDNA molecule, respectively. The primers are numbered according to the location of their 5'-ends in the rCRS [26, 27], except for –21M13L00182, M13revH00619, L16196 and L00318, which are numbered according to the location of their 3'-ends in the rCRS

| Amplification primer | Nucleotide sequence |
|---|---|
| L15900 | 5' TAC ACC AGT CTT GTA AAC C 3' |
| H00599 | 5' TTG AGG AGG TAA GCT ACA TAA 3' |
| –21M13L00182 | 5' GTA AAA CGA CGG CCA GTG ACG CAC CTA CGT TCA ATA TTA C 3' |
| M13revH00619 | 5' CGG AAA CAG CTA TGA CCA TGG GTG ATG TGA GCC CCG TCT AA 3' |
| **Sequencing primer** | **Nucleotide sequence** |
| L15971 | 5' TTA ACT CCA CCA TTA GCA CC 3' |
| L15989 | 5' CCC AAA GCT AAG ATT CTA AT 3' |
| L16196 | 5' CCC CCC CCC CAT G 3' |
| L16268 | 5' CAC TAG GAT ACC AAC AAA CC 3' |
| L00015 | 5' CAC CCT ATT AAC CAC TCA CG 3' |
| L00318 | 5' CCC CCC CCC CCC GCT 3' |
| L00403 | 5' TCT TTT GGC GGT ATG CAC TTT 3' |
| H16410 | 5' GAG GAT GGT GGT CAA GGG AC 3' |
| H00159 | 5' AAA TAA TAG GAT GAG GCA GGA ATC 3' |
| H00484 | 5' TGA GAT TAG TAG TAT GGG AG 3' |
| –21M13 | 5' GTA AAA CGA CGG CCA GTG A 3' |
| M13rev | 5' GGA AAC AGC TAT GAC CAT G 3' |

H00159, and H00484. Samples showing length heteroplasmy in HV1 (15%) were alternatively sequenced with primers L15989 and L16196 [24]. Samples showing length heteroplasmy in HV2 were sequenced with L00318. Samples with an $(AC)_n$ repeat polymorphism located in HV3 were reamplified using the primer pair –21M13L00182/H00619M13Rev [25]. These templates were sequenced with the universal primer (M13Rev). The primers are numbered according to the location of their 5'-ends in the revised Cambridge Reference Sequence (rCRS) [26, 27], except for L16169 and L00182, which are numbered according to the location of their 3'-ends in the rCRS, and "L" and "H" designate the light and the heavy strands of the mtDNA molecule, respectively. Primers were synthesized by Eurogentec (Seraing, Belgium) and Metabion (Martinsried, Germany). Sequencing reaction products were purified from residual dye terminators using Sephadex G-50 Fine (Amersham, Buckinghamshire, UK) and Multiscreen filter plates (Millipore). Electrophoretic separation was carried out on an ABI3700 capillary sequencer. In order to confirm heteroplasmic point substitutions, the samples concerned from both mother and child were repeated with alternative primers.

### Data analysis

Sequences were aligned and compared to the rCRS using Sequencher (Version 4.1.4Fb4, GeneCodes, Ann Arbor, MI), following international guidelines for mtDNA typing [28, 29, 30]. As an internal control for sequence quality, sequence evaluation was replicated by independent laboratory personnel for all mothers, for a random sample of the children, and for all heteroplasmic individuals sequenced in this study. In all cases, the sequencing results matched exactly between evaluations.

### Characterization of heteroplasmy

A nucleotide position was considered heteroplasmic if a secondary peak of more than about 10% of the height of the primary peak was present. This needed to be confirmed with at least one forward and one reverse sequencing reaction (if reverse sequencing reactions failed to give high quality electropherograms, two forward reactions were performed with alternative primers). Heteroplasmic positions were designated using the IUPAC nomenclature. In the case of homopolymeric cytosine tracts, where C-insertions often lead to length heteroplasmy, we determined the number of cytosines by counting the number of "dominant Cs", i.e. cytosines clearly above the background and sequence carry-over level. A C-insertion was established when all sequence reactions confirmed the result. Point heteroplasmic mixtures were characterized by comparing the heights of the two contributing peaks. The proportion of the secondary peak was estimated by measuring the heights of the primary and secondary peaks of mixed basecalls as depicted in the electropherograms (Sequence Navigator, AB). The numerical proportions of the Cartesian coordinates from the tops of upper and lower basecalls were inferred from all electropherograms from a point heteroplasmic sample. The obtained numerical proportions were again averaged to give a final estimate of the amount of the minor component in a heteroplasmic mixture.

### Sample quality values

All mother-child pairs, in which at least one of the two generations showed point heteroplasmy in the mtDNA CR, were reanalyzed using the sequence analysis and alignment software SeqScape (Version 2.0, AB). The software calculates a quality value (QV), which is a per-base estimate of the basecaller accuracy. Per-base QVs are calibrated on a scale corresponding to $QV=-10\log_{10}(Pe)$, where $Pe$ is the probability of error of the base call. The range of a QV is 1–50, with 1 being low confidence and 50 being high confidence. Mixed basecalls yield lower position QVs than pure basecalls. The software allows the user to search specifically for positions with low quality values with the tab stop function. Beforehand, the user specifies the range of low, medium and high quality QVs. For every point heteroplasmic sample (and samples, which were maternally related to point heteroplasmic samples), a segment score, which is the average quality value of the bases in the clear range of a sample, was calculated for the segment harboring the point heteroplasmic position (16024–16569 or 1–576) by the SeqScape software.

The existence of a correlation between the heights of the secondary peaks in mixed basecalls and the corresponding QVs at heteroplasmic positions (as the mean value of the QVs of aligned sequences at the heteroplasmic position) was tested with the software package SPSS (Version 11.0.1, SPSS Inc., Chicago, Il). The calculation was based on the quality values of all substitution polymorphisms relative to the rCRS in the segment harboring point heteroplasmy of heteroplasmic samples and their maternal relatives.

## Results and discussion

The entire mitochondrial DNA control region (1122 bp) of 270 blood samples (135 mother-child pairs) was determined by direct sequencing for population databasing purposes (the population data will be published online in the near future). Overall, 135 generational events were examined. No mutations in the sense of nucleotide substitutions were observed but 20 mother-offspring pairs demonstrated more than 1 mtDNA haplotype (heteroplasmy). In 13 families, the observed heteroplasmic substitutions led to differences in the haplotypes between mother and offspring (Table 2). In 5 of the 13 lineages (6 individual positions) heteroplasmic point mutations were the source of the difference, in 7 lineages C-insertions and in 1 family an AC-deletion in HV3 were the reason for the divergent haplotypes. In two cases, point heteroplasmy of the maternal sequence was resolved to homoplasmy in the corresponding sequence of the child. The other families demonstrated heteroplasmy in the children's sequence, while the corresponding maternal positions were homoplasmic. While we did not see instances of apparent substitution, we saw shifts in major sequence types in 9.6% of all studied meiotic events. These results are well in agreement with the occurrence of rapid base changes from mother to offspring as postulated by the bottleneck theory [16, 17, 18, 19].

### Point heteroplasmy

Single point heteroplasmic positions were found in 10 mothers, but only 8 of them passed on this feature to their offspring, at least at the level of detection. For one child, the electropherograms were suggestive that heteroplasmy might be present at very low levels, but firm conclusions could not be drawn from the sequence data. On the other hand, three children carried point heteroplasmies, which were not visible in the corresponding sequences of their mothers (Table 2). Overall, we observed five differences between mothers and their offspring with respect to detectable point heteroplasmic mixtures, four out of these five sites were pyrimidine transitions and one site was a purine transition. In two children, point heteroplasmy was observed and confirmed at two different positions (16519Y and 154Y in one child; 152Y and 194Y in the other child)

**Table 2** Profiles of 20 mother and child samples, sequenced for the entire mtDNA control region

| rCRS | 16093 T | 16216 A | 16220 A | 16232 C | 16301 C | 16311 T | 16355 C | 16362 T | 16519 T | 152 T | 154 T | 194 C | 234 A | 309.1 – | 309.2 – | 523 A | 524 C | 573.5 – |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| m1_B3 | | | | | | | Y (C[a]) | | | | | | | | | | | |
| c1_B3 | | | | | | | C | | | | | | | | | | | |
| m1_B5 | | | | | | | | | | Y (T[a]) | | | | | | | | |
| c1_B5 | | | | | | | | | | Y (C[a]) | | | | | | | | |
| m1_C6 | Y (C[a]) | | | | | | | | | | | | | | | | | |
| c1_C6 | Y (C[a]) | | | | | | | | | | | | | | | | | |
| m1_E5 | | | | | | | | | | | | | | | – | | | |
| c1_E5 | | | | | | | | | | | | | | | C | | | |
| m1_H1 | | | | | | | | | | | | | | – | | | | |
| c1_H1 | | | | | | | | | | | | | | C | | | | |
| m1_H3 | | | | | | | | | | | | | | | – | | | |
| c1_H3 | | | | | | | | | | | | | | | C | | | |
| m2_A4 | | | | | | | | | T | | T | | | | | | | |
| c2_A4 | | | | | | | | | Y (T[a]) | | Y (C[a]) | | | | | | | |
| m2_B1 | | | | | | | | | | | | | | – | | | | |
| c2_B1 | | | | | | | | | | | | | | C | | | | |
| m2_B4 | | | | | | | | | | | | | | | | | | – |
| c2_B4 | | | | | | | | | | | | | | | | | | C |
| m2_E4 | | | | | | | | | | | | | | | – | | | |
| c2_E4 | | | | | | | | | | | | | | | C | | | |
| m2_E5 | | | | S (C[a]) | | | | | | | | | | | | | | |
| c2_E5 | | | | S (C[a]) | | | | | | | | | | | | | | |
| m2_F1 | | | | | | C | | | | | | | | | | | | |
| c2_F1 | | | | | | Y (C[a]) | | | | | | | | | | | | |
| m2_H1 | | | | | | | | | | Y (C[a]) | T | | | | | | | |
| c2_H1 | | | | | | | | | | Y (T[a]) | Y (T[a]) | | | | | | | |
| m2_H5 | | | R (G[a]) | | | | | | | | | | | | | | | |
| c2_H5 | | | R (G[a]) | | | | | | | | | | | | | | | |
| m3_B2 | | | | | | | | | | | | | R (A[a]) | | | | | |
| c3_B2 | | | | | | | | | | | | | R (G[a]) | | | | | |
| m3_B5 | | | | | Y (C[a]) | | | | | | | | | | | | | |
| c3_B5 | | | | | Y (C[a]) | | | | | | | | | | | | | |
| m3_C2 | | | | | | | | | | | | | | | | A | C | |
| c3_C2 | | | | | | | | | | | | | | | | del | del | |
| m3_D5 | | | | | | | | | | | | | | | – | | | |
| c3_D5 | | | | | | | | | | | | | | | C | | | |
| m3_G6 | | R (A[a]) | | | | | | | | | | | | | | | | |
| c3_G6 | | A | | | | | | | | | | | | | | | | |
| m3_H3 | | | | | | | | Y (C[a]) | | | | | | | | | | |
| c3_H3 | | | | | | | | Y (C[a]) | | | | | | | | | | |

The exact number of insertions in the poly-C tract was unclear, therefore at position 309.1 and 309.2 the dominant number of C-insertions is given.
[a]Indicates the nucleotide which accounts for the bigger proportion in the mixture.

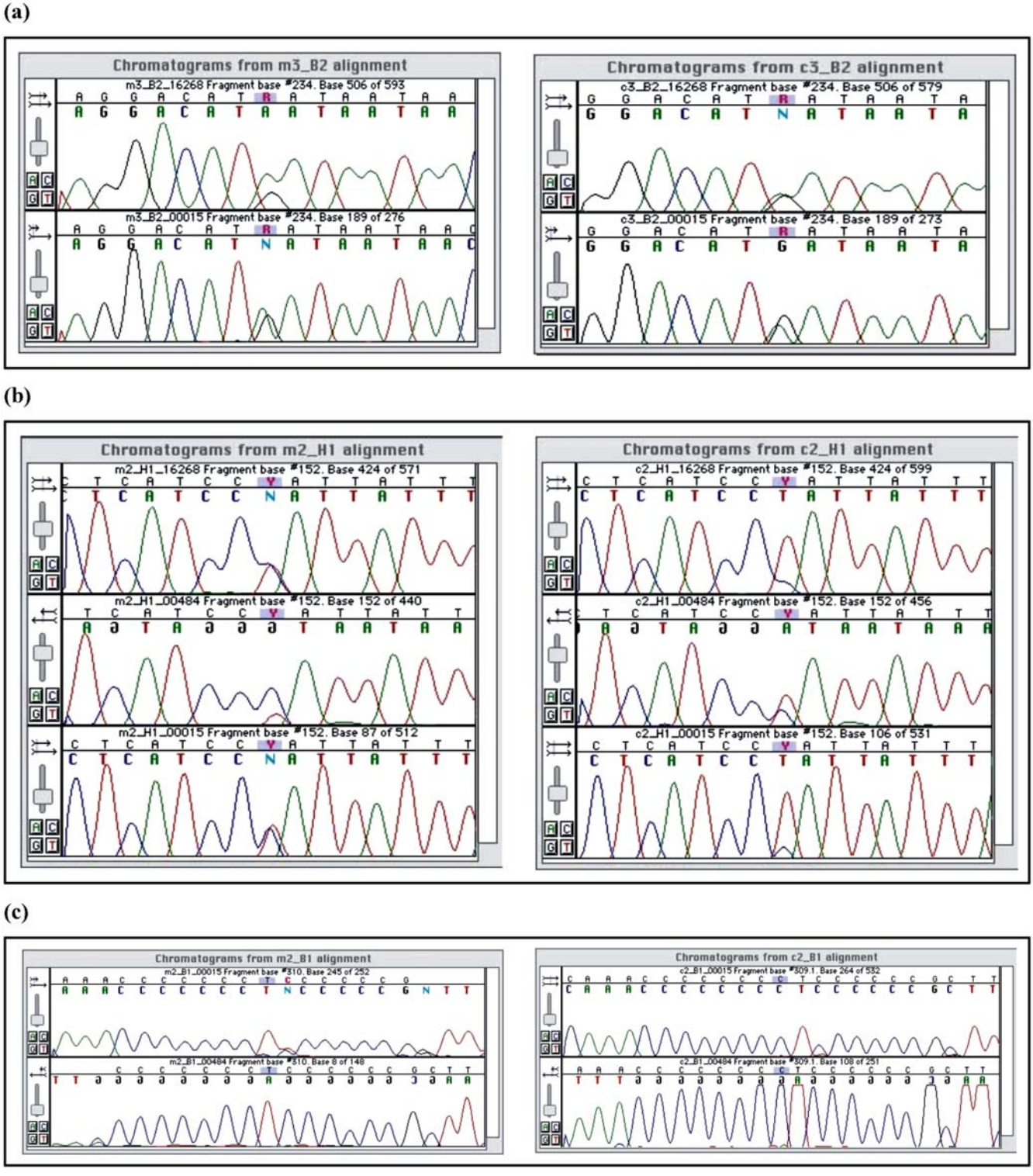


**Fig. 2a–c** Electropherograms of three mother-child pairs showing heteroplasmic substitutions. The mothers' profiles are depicted in the *left panels*; the children's profiles are depicted in the *right panels*. **a** Different mixture ratio of purines at heteroplasmic position 234 in HV2 between mother and child. **b** Different mixture ratio of pyrimidines at heteroplasmic hotspot 152 in HV2 between mother and child. **c** The child shows an additional C-insertion in the HV2 C-stretch compared to the mother

– a state known as "triplasmy" [31, 32]. Among the 20 heteroplasmic individuals, point heteroplasmy was found at 13 different positions. One position (152) showed heteroplasmy in four individuals (two mother-child pairs). Additional observations of point heteroplasmy were made at positions 16093, 16216, 16220, 16232, 16301, 16311, 16355,

16362, 16519, 152, 154, 194 and 234. Hence, point heteroplasmy occurred in 2.3% of all sequenced base positions in HV1, but only in 1.5% of all base positions in HV2. In two mother-child pairs, the peak height proportion of both nucleotides at the point heteroplasmic positions showed differences, the dominant type switched between the two generations (Table 2 and Fig. 2a,b).

## Length heteroplasmy

Of the mother-child pairs, eight showed differences with respect to length heteroplasmy. In two cases, the child showed an additional cytosine insertion in the HV2 polycytosine region (C-stretch) between positions 303–309. The 309.1C insertion is a common variant, shared by 56% of individuals in our database of 270 CR sequences from unrelated persons. Most individuals are heteroplasmic at low levels for length variants in this region. In four cases, the mother carried the C-insertion at 309.1 C, but the corresponding children showed an additional insertion to 309.2 C (Fig. 2c). In one case, the child showed a deletion in HV3 [25] at positions 523A and 524C, but the corresponding mother did not. One woman had 4 C-insertions at position 573, while the child showed 5 insertions.

## Statistics

Within 135 unrelated women, 10 women showed point heteroplasmy, corresponding to an incidence of 7.4% (95% CI 3.0%–11.8%) in the population sample. Interestingly, the same number of point heteroplasmic individuals was found in the offspring population, even though only 80% of point heteroplasmic mothers transmitted this feature to their children. Out of 135 unrelated women, 73 carried the polymorphism 309.1C corresponding to an occurrence of 54.1% (95% CI 45.6%–62.6%), 75 out of 135 children however, showed the 309.1C insertion in the HV2 C-stretch, indicating a frequency of 55.6% (95% CI 47.0%–64.0%).

**Table 3** Quality values of point-heteroplasmic mother-offspring samples

| Sample | Segment | Segment score | Sequences[a] | Position | Position QV | Secondary peak[b] | Proportion[c] | Automatic IUB code[d] | Found with tab stop[e] |
|---|---|---|---|---|---|---|---|---|---|
| m1B3 | 16024–16569 | 46 | 4 | 16355Y | 34 | T | 0.15 | No | Yes |
| c1B3 | 16024–16569 | 45 | 4 | 16355C (rCRS) | 50 | – | 0.00 | – | – |
| m1B5 | 1–576 | 47 | 5 | 152Y | 33 | C | 0.48 | Yes | Yes |
| c1B5 | 1–576 | 40 | 4 | 152Y | 24 | T | 0.13 | No | Yes |
| m1C6 | 16024–16569 | 45 | 4 | 16093Y | 29 | T | 0.66 | Yes | Yes |
| c1C6 | 16024–16569 | 49 | 5 | 16093Y | 24 | T | 0.72 | Yes | Yes |
| m2A4 | 16024–16569 | 42 | 3 | 16519T (rCRS) | 42 | – | 0.00 | – | – |
| c2A4 | 16024–16569 | 42 | 3 | 16519Y | 29 | C | 0.09 | No | No |
| m2A4 | 1–576 | 44 | 5 | 154T (rCRS) | 46 | – | 0.00 | – | – |
| c2A4 | 1–576 | 47 | 6 | 154Y | 29 | T | 0.86 | Yes | Yes |
| m2E5 | 16024–16569 | 46 | 4 | 16232S | 26 | G | 0.52 | Yes | Yes |
| c2E5 | 16024–16569 | 45 | 4 | 16232S | 3 | G | 0.49 | Yes | Yes |
| m2F1 | 16024–16569 | 46 | 4 | 16311C | 45 | – | 0.00 | – | – |
| c2F1 | 16024–16569 | 47 | 4 | 16311Y | 17 | T | 0.20 | No | Yes |
| m2H1 | 1–576 | 43 | 5 | 152Y | 26 | T | 0.73 | Yes | Yes |
| c2H1 | 1–576 | 48 | 6 | 152Y | 32 | C | 0.36 | Yes | Yes |
| m2H1 | 1–576 | 43 | 5 | 194T | 41 | – | 0.00 | – | – |
| c2H1 | 1–576 | 48 | 6 | 194Y | 17 | C | 0.19 | No | Yes |
| m2H5 | 16024–16569 | 40 | 6 | 16220R | 23 | A | 0.57 | Yes | Yes |
| c2H5 | 16024–16569 | 42 | 6 | 16220R | 28 | A | 0.71 | Yes | Yes |
| m3B2 | 1–576 | 40 | 6 | 234R | 25 | G | 0.66 | Yes | Yes |
| c3B2 | 1–576 | 40 | 6 | 234R | 28 | A | 0.67 | Yes | Yes |
| m3B5 | 16024–16569 | 46 | 4 | 16301Y | 28 | T | 0.15 | No | Yes |
| c3B5 | 16024–16569 | 45 | 4 | 16301Y | 40 | T | 0.13 | No | No |
| m3G6 | 16024–16569 | 46 | 4 | 16216R | 25 | G | 0.31 | Yes | Yes |
| c3G6 | 16024–16569 | 46 | 4 | 16216A (rCRS) | 43 | – | 0.00 | – | – |
| m3H3 | 16024–16569 | 46 | 4 | 16362Y | 29 | T | 0.82 | Yes | Yes |
| c3H3 | 16024–16569 | 47 | 4 | 16362Y | 30 | T | 0.35 | Yes | Yes |

*rCRS* Indicates that the corresponding base is identical to the rCRS base at that position.

[a]Number of different sequence reactions for the individual segment.

[b]Base that accounts for the minor component in a heteroplasmic mixture.

[c]Estimated proportion of the height of the secondary peak on the height of the primary peak.

[d]Cases in which the heteroplasmic mixture was automatically detected by the sequence analysis software SeqScape. The level for calling IUB-codes was set to 10%.

[e]Case where the DNA analyst was alerted to check a certain position because of its low QV.

An increase of the frequency of a C-insertion was also observed at position 309.2C. While only 18 mothers or 13.3% (95% CI 7.5%–19.1%) were carriers of this feature, the insertion polymorphism 309.2C was found in 22 children or 16.3% (95% CI 10.0%–22.6%). The same phenomenon was observed in the HV3 $(AC)_n$ short tandem repeat. While 7 out of 135 women showed a deletion of one AC with regard to the rCRS (5.2%; 95% CI 1.4%–9.0%), the number of infants carrying an AC deletion was increased by one individual to a frequency of 5.9% (95% CI 1.9%–10.0%).

## Sequence quality values

For all point-heteroplasmic mother-offspring samples, QVs were computed for the region which contained the heteroplasmic position. The segment score is the average quality value of the bases in the range of the sample that harbors the heteroplasmy (this was either 16024–16569 or 1–576, respectively) (Table 3). The position QV is the average of the per-base estimates of the basecaller accuracy of aligned sequences at the heteroplasmic position. Overall, point heteroplasmic samples (and samples, which were maternally related to point heteroplasmic samples, *n*=40) showed an average segment score of 44.7, corresponding to a mean error probability of 0.003% for every base position in the consensus sequence. Additionally, for all individual positions, where point heteroplasmy was detected in either the mother or the child sample (or in both samples), the QV value of this position was determined (Table 3). Furthermore, for point heteroplasmic positions, the proportional height of the secondary peak to the primary peak was estimated by comparison of Cartesian Y-coordinates with Sequence Navigator (Table 3). This information was compared with the QVs obtained for these positions in order to evaluate the usefulness of QVs for automated mixture detection, bearing in mind that the results would reflect a chemistry-biased value rather than the actual distribution of the different mtDNA populations. However, at least the procedure of comparing peak heights visually to each other – which is the current method commonly used – could be standardized and automated. The results are summarized in Table 4 and Fig. 3. Obviously, mixed basecalls yielded lower QVs than clear basecalls; in fact, the position QV was moderately, but significantly negatively correlated with the estimated proportional height of the secondary peak (Pearson correlation coefficient of –0.754; $p_\alpha$ (double-sided): 0.000; *n*=106).

**Table 4** Descriptive statistics of the influence of the estimated proportional height of the secondary peak on the position QV

| Proportional height of secondary peak | N | Position QV (mean) |
|---|---|---|
| 0.00 | 84 | 46.0 (s.d. 4.0) |
| 0.09–0.31 | 7 | 28.1 (s.d. 7.4) |
| 0.35–0.57 | 7 | 23.4 (s.d. 10.6) |
| 0.66–0.86 | 8 | 27.3 (s.d. 2.0) |

*0.00* Refers to all substitution polymorphisms relative to the rCRS in the region harboring a point heteroplasmy of heteroplasmic samples and their maternal relatives where no point heteroplasmy was detected by direct sequencing.
*s.d.* Standard deviation.

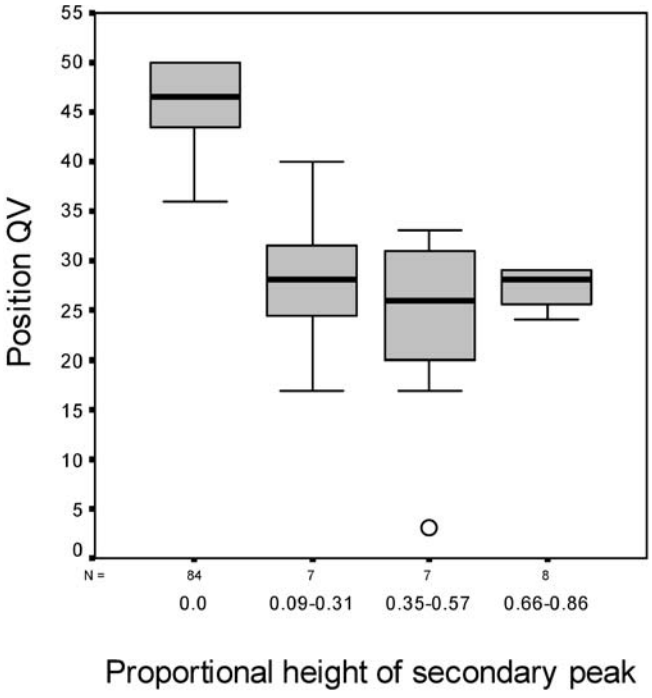


**Fig. 3** Position quality values with respect to the estimated proportional height of the secondary peak. The horizontal line in the middle shows the median of the sample. The top and bottom of the boxes show the 75th and 25th percentiles, respectively. In this graph, the top and bottom of the whiskers show the maximum and minimum values. One outlier was observed (*circle*)

SeqScape allows the user to apply a mixed base identification algorithm. For this study, the algorithm was tuned to call IUB codes if the minor peak at a particular base position was equal to/greater than 10% of the major peak. Indeed, 15 out of 22 point heteroplasmies were automatically detected by the software (Table 3). These heteroplasmic positions, however, were mixtures where the minor component accounted for at least 30% of the height of the upper peak. The remaining heteroplasmic positions, all with minor components of 9–20% of major peak heights, were designated with the dominant basecall. Nonetheless, if the specimen was searched for low quality basecalls with the tab stop function – a method used to quickly screen through polymorphisms, 20 out of 22 point heteroplasmies could have been detected. Only two point heteroplasmies (with proportional heights of the secondary peak of 9% and 13%, respectively) could not be detected with the search algorithms of the software.

In summary, in more than 90% of all cases it was possible to detect the point heteroplasmic state by the basecall and search algorithms provided by the analysis software. Thus, the implementation of quality values can be a very useful tool for semi-automated detection of point

heteroplasmies, provided that the overall sequence quality is high. For low quality sequences, in this case for sequences with a very high sequence background, the detection of low level heteroplasmy is nearly impossible [33].

However, a visual check of the entire sequence remains to be of great importance in order to address low level point heteroplasmy and in order to confirm the programme-based recommendation of the haplotype. However, the identification of point heteroplasmy with the aid of quality values could be facilitated and could hence be standardized for forensic mtDNA sequence analysis.

## Comparison to previous studies

In recent years, two studies on the heredity of mitochondrial heteroplasmy have contributed much to the understanding of mtDNA sequence variation in forensic investigations [21, 22]. The aim of this study was to extend the body of acquired knowledge by analyzing the entire mitochondrial control region, which offered the possibility to target highly variable polymorphisms outside HV1/HV2, for example T16519C, the HV3 $(AC)_n$ STR, and the C-stretch on the 3'-end of the CR. As a matter of fact, 3 out of 13 discordant mother-offspring haplotypes were caused by heteroplasmic substitutions in these regions. Although accurate estimates of the proportion of minor variants in C-stretches could not be drawn from direct sequence analyses, the sample size in the current study enabled an overview of the frequency of intergenerational heteroplasmic substitutions and thus allowed further inspection of the mechanisms of the assumed mitochondrial genetic bottleneck.

**Acknowledgements** The authors would like to thank Cordula Eichmann and Anna König (Institute of Legal Medicine, University of Innsbruck, Austria) for valuable discussion.

## References

1. Gocke CD, Benko FA, Rogan PK (1998) Transmission of mitochondrial DNA heteroplasmy in normal pedigrees. Hum Genet 102:182–186
2. Chinnery PF, Thorburn DR, Samuels DC et al. (2000) The inheritance of mitochondrial DNA heteroplasmy: random drift, selection or both? Trends Genet 16:500–505
3. 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
4. Calloway CD, Reynolds RL, Herrin GL Jr, Anderson WW (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
5. Hutchison CA III, Newbold JE, Potter SS, Edgell MH (1974) Maternal inheritance of mammalian mitochondrial DNA. Nature 251:536–538
6. Hagelberg E (2003) Recombination or mutation rate heterogeneity? Implications for mitochondrial Eve. Trends Genet 19: 84–90
7. Awadalla P, Eyre-Walker A, Smith JM (1999) Linkage disequilibrium and recombination in hominid mitochondrial DNA. Science 286:2524–2525
8. Eyre-Walker A (2000) Do mitochondria recombine in humans? Philos Trans R Soc Lond B Biol Sci 355:1573–1580
9. Strauss E (1999) Human genetics. mtDNA shows signs of paternal influence. Science 286:2436
10. Hey J (2000) Human mitochondrial DNA recombination: can it be true? Trends Ecol Evol 15:181–182
11. Parsons TJ, Irwin JA (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288:1931
12. Hedrick P, Kumar S (2001) Mutation and linkage disequilibrium in human mtDNA. Eur J Hum Genet 9:969–972
13. Jorde LB, Bamshad M (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288:1931
14. Kivisild T, Villems R (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288:1931
15. Kumar S, Hedrick P, Dowling T, Stoneking M (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288:1931
16. Shoubridge EA (2000) Mitochondrial DNA segregation in the developing embryo. Hum Reprod 15 [Suppl 2]:229–234
17. Jansen RP (2000) Germline passage of mitochondria: quantitative considerations and possible embryological sequelae. Hum Reprod 15 [Suppl 2]:112–128
18. Bianchi NO, Bianchi MS, Richard SM (2001) Mitochondrial genome instability in human cancers. Mutat Res 488:9–23
19. Howell N, Chinnery PF, Ghosh SS, Fahy E, Turnbull DM (2000) Transmission of the human mitochondrial genome. Hum Reprod 15 [Suppl 2]:235–245
20. Parsons TJ, Muniec DS, Sullivan K et al. (1997) A high observed substitution rate in the human mitochondrial DNA control region. Nat Genet 15:363–368
21. Huhne J, Pfeiffer H, Brinkmann B (1999) Heteroplasmic substitutions in the mitochondrial DNA control region in mother and child samples. Int J Legal Med 112:27–30
22. Lutz S, Weisser HJ, Heizmann J, Pollak S (2000) Mitochondrial heteroplasmy among maternally related individuals. Int J Legal Med 113:155–161
23. Walsh PS, Metzger DA, Higuchi R (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques 10:506–513
24. Szibor R, Michael M (1999) Correct mitochondrial L-strand sequencing after C-stretches. Int J Legal Med 112:348–349
25. 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
26. Anderson S, Bankier AT, Barrell BG et al. (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465
27. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147
28. Bär W, Brinkmann B, Budowle B et al. (2000) DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. Int J Legal Med 113:193–196
29. Carracedo A, Bär W, Lincoln P et al. (2000) DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. Forensic Sci Int 110:79–85
30. Wilson MR, Allard MW, Monson K, Miller KW, Budowle B (2002) Recommendations for consistent treatment of length variants in the human mitochondrial DNA control region. Forensic Sci Int 129:35–42
31. Bidooki SK, Johnson MA, Chrzanowska-Lightowlers Z, Bindoff LA, Lightowlers RN (1997) Intracellular mitochondrial triplasmy in a patient with two heteroplasmic base changes. Am J Hum Genet 60:1430–1438
32. Tully LA, Parsons TJ, Steighner RJ, Holland MM, Marino MA, Prenger VL (2000) A sensitive denaturing gradient-gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region 1 of the human mtDNA control region. Am J Hum Genet 67:432–443
33. Brandstätter A, Parson W (2003) Mitochondrial DNA heteroplasmy or artefacts – a matter of the amplification strategy? Int J Legal Med 117:180–184