Characterization of mtDNA SNP typing using quantitative rtPCR for forensic purposes

Harald Niederstätter[a], Michael D. Coble[b,c], Thomas J. Parsons[b] & Walther Parson[a]

[A]Institute of Legal Medicine, Innsbruck Medical University, Austria
[b]Armed Forces DNA Identification Laboratory, Armed Forces Institute of Pathology, Rockville, MD, USA
[C]Biotechnology Division, National Institute of Standards and Technology, Gaithersburg, MD, USA

Introduction

The major limitation of forensic mitochondrial DNA (mtDNA) typing is the low power of discrimination that is obtained when common haplotypes are present. Current mtDNA testing typically targets one or two hypervariable regions in the non-coding control region of the mitochondrial genome by sequencing a total of approx. 600 base pairs (~3.6% of the entire mitochondrial genome). It is now increasingly recognized that assays targeting single nucleotide polymorphisms (SNPs) are well suited for gaining additional information in mtDNA testing.

We chose to investigate the forensic applicability of real-time PCR (rtPCR) using a quantitative allele-discriminatory duplex TaqMan assay targeted to the highly informative mtSNP 16519T/C for several reasons:

1) Its large linear dynamic range allows the analysis of a broad range of specimens under identical technical conditions, regardless of DNA concentration.

2) The short amplicon lengths obtained enable the analysis of even severely fragmented DNA.

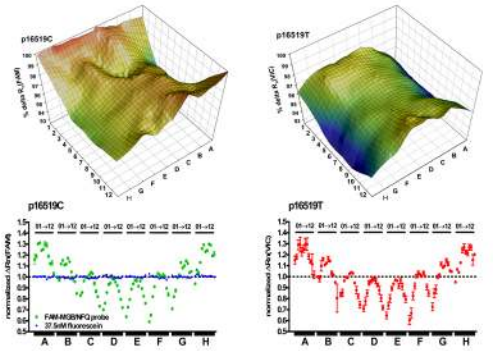
3) Its homogeneous format largely reduces the risk of sample mix-up and cross-contamination of samples with PCR products, and makes SNP typing easy to automatize.

4) The quantitative nature of the TaqMan assay makes it unique among recent SNP typing methods, providing the potentially very useful means for distinguishing between authentic signal and possible contamination based on measurable values.

5) The multicolor capability of rtPCR instruments enables the simultaneous interrogation of both base states of SNPs, which is particularly important for mixtures/heteroplasmy detection and allele proportion assessment.

Materials and Methods

We performed a study on 135 paternity trios with known control region sequences using differently labeled allele- specific TaqMan hybridization probes with an attached minor groove binding moiety and a PE Applied Biosystems 7700 sequence detection system. Cloned and sequence verified alleles p16519T and p16519C were quantified and used to produce calibration curves and defined mixtures with known copy numbers. For three heteroplasmic and two homoplasmic samples – as determined by rtPCR – the base state at nucleotide position 16519 was validated by cloning the PCR amplified control regions and subsequent rtPCR typing of approx. 300 colonies per sample. Automated data analysis was performed by pasting an MS Excel template into the results file. Absolute quantification was based on C_T values with reference to the parameters of pDNA calibration curves. For allele calling and mixture ratio assessment %ΔRn values [e.g. %ΔRn(FAM) = ΔRn(FAM) / (ΔRn(FAM) + ΔRn(VIC))] were used.



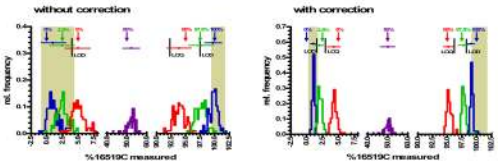


Figure 2. Heterogeneity of %ΔRn **(TOP)** and normalized ΔRn **(MIDDLE)** values caused by the real-time PCR instrument. For both targets (100% p16519T or C) 4 independent amplifications of 96 replicates were performed. The obtained data were used to calculate corrected %ΔRn values. **BOTTOM:** Frequency distributions obtained with uncorrected and corrected %ΔRn values for defined p16519T and C mixtures (n = 96 replicates each; 50,000 plasmids/reaction). Average measured contributions of p16519C to the pDNA mixtures and the 95% prediction intervals are shown in graphs as dots and horizontal lines. **Shaded areas** depict the range of values found in the concordance study for the 402 homoplasmic gDNA samples. **LOD/LOQ**: Limit of detection/ quantification

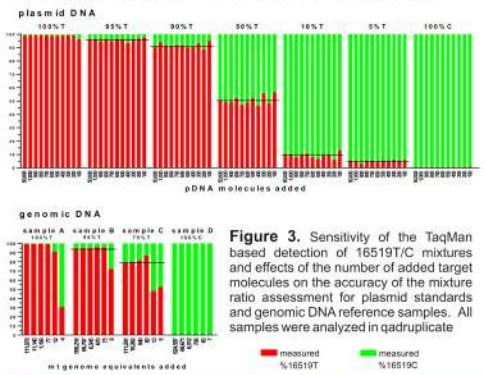


Figure 3. Sensitivity of the TaqMan based detection of 16519T/C mixtures and effects of the number of added target molecules on the accuracy of the mixture ratio assessment for plasmid standards and genomic DNA reference samples. All samples were analyzed in qadruplicate

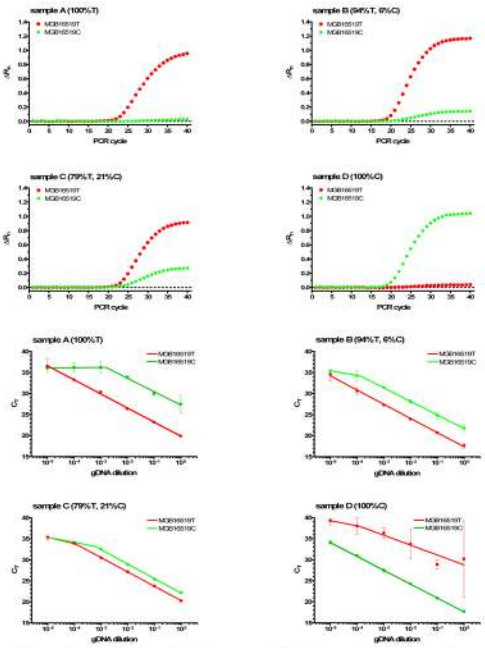


Figure 1. Amplification plots (**upper panels**) and linear dynamic range (**lower panels**) of homoplasmic (A, D) and heteroplasmic (B, C) genomic DNA reference samples

Results

The results of a study on 405 persons with known mitochondrial control region sequences showed that both alleles of the mtSNP 16519 can be reliably and simultaneously typed with the TaqMan approach. For both base states the linear dynamic range covered at least 5 orders of magnitude with a sensitivity of approx. 10 dsDNA template molecules and an apparent single cycle PCR efficiency close to 100% during the exponential phase of the reaction (Fig. 1, Table 1). As differences between the PCR efficiencies obtained for circular plasmid DNA standards and genomic DNA were statistically insignificant (p_{α}>0.05), pDNA standard curves were used for absolute quantification.

The minor component in defined mixtures of p16519T and p16519C could be detected and quantified reliably down to the 5% level for both alleles without a need for replicate reactions when corrected %ΔRn values were used for data analysis (Fig.2). Heteroplasmy detection at the 2.5% level was possible when samples were run in replicates (see LOD and 95% prediction intervals in Fig. 2, bottom panels).

The sensitivity for accurate heteroplasmy detection and mixture ratio assessment at several levels (0%, 5%, 10% and 50% minor component) was approx. 100 - 200 copies (Fig. 3).

The mixture ratios found for 3 heteroplasmic and 2 homoplasmic gDNA samples were concordant with the results obtained by typing approx. 300 cloned mitochondrial control regions per sample and those derived from the ratios of the peak heights in sequencing electropherograms (Table 2, Fig. 4). Contrary to rtPCR, however, pose neither cloning nor Phred quality value assisted analysis of sequencing traces (Fig. 4) feasible methods for fully automated detection and quantification of low level mixtures/point-heteroplasmy.

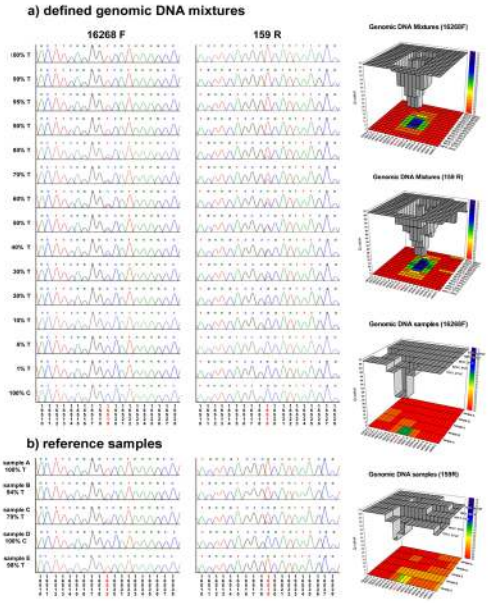


Figure 4. Sequencing electropherograms obtained for defined mixtures of two genomic DNA samples, two homoplasmic and three heteroplasmic reference samples, using BigDye terminator chemistry (v2.0, Applied Biosystems). Laser induced fluorescence capillary electrophoresis was performed on an ABI 3700 DNA analyzer. Quality values for forward and reverse strand sequencing traces were computed with MacPhred (Phred version 0.020425.c)

pDNA vs. genomic DNA	standard curves		sample A	sample B		sample C		sample D
	100% p16519T	100% p16519C	100% T	94% T	6% C	79% T	21% C	100% C
slope ± SE	-3.415 ± 0.060	-3.214 ± 0.080	-3.355 ± 0.097	-3.369 ± 0.109	-3.226 ± 0.195	-3.404 ± 0.084	-3.462 ± 0.052	-3.296 ± 0.043
Y-intercept ± SE	37.12 ± 0.225	36.38 ± 0.280	36.72 ± 0.203	36.60 ± 0.389	36.39 ± 0.529	36.98 ± 0.139	37.41 ± 0.176	36.77 ± 0.222
r^2	0.994	0.987	0.992	0.990	0.990	0.998	0.999	0.998
PCR efficiency	92.26%	104.71%	98.64%	98.07%	104.17%	96.69%	94.47%	101.09%
lin. dyn. range [copies]	10 - 1,000,000	1 - 1,000,000	10 – (109,689)	4 – (529,811)	7 – (35,686)	7 – (83,636)	13 – (26,908)	7 – (613,503)

Table 1. Comparison of the plasmid DNA standard curves for 16519T and 16519C with homo/heteroplasmic reference gDNA samples . The slopes and Y-intercepts were obtained by linear regression analysis, plotting C_T values (specific signal) versus the $\log_{10}$ of the dsDNA template molecule number added. The apparent single-cycle PCR efficiencies during the exponential phase were calculated from the slopes of the regression lines as $E = 100x[10^{(-1/slope)} - 1]$. For genomic DNA samples (**A** - **D**) the calculated "high-end" copy numbers were set in parentheses because they do not necessarily represent the upper limit of the linear dynamic range

TaqMan vs. Cloning and Sequencing		TaqMan		Cloning		Sequencing	
		%16519C	95% CI	%16519C	95% CI	%16519C	95% CI
sample A	avg	0.91%	(0.73% - 1.10%)	0.35%	(0.01% - 1.95%)	1.51%	(0.35% - 2.68%)
	n	9		(1/284)			
sample B	avg	6.20%	(5.62% - 6.78%)	4.18%	(2.18% - 7.19%)	6.06%	(0.88% - 11.23%)
	n	18		(12/287)			
sample C	avg	21.43%	(20.68% - 22.18%)	17.12%	(13.23% - 21.6%)	17.25%	(1.78% - 32.71%)
	n	18		(57/333)			
sample D	avg	99.84%	(99.77% - 99.91%)	100.00%	(99.04% - 100%)	99.84%	(98.37% - 100%)
	n	9		(384/384)			
sample E	avg	1.95%	(1.81% - 2.09%)	2.08%	(0.77% - 4.48%)	0.00%	(-)
	n	10		(6/288)			

Table 2. Comparison of the SNP typing results obtained with real-time PCR (corrected %ΔRn values used), the typing of individual bacterial colonies containing cloned mitochondrial control regions, and the ratio of the peak heights in the sequencing electropherograms. For the TaqMan and the sequencing approach calibration curves were constructed by mixing p16519T and p16519C at defined ratios (50,000 dsDNA molecules in total). Sample E: mother of B

Acknowledgements: Part of this work was supported by National Institute of Justice grant 2000-1J-CX-K010 to T.J.P. The opinions and assertions contained herein are solely those of the authors and are not to be construed as official or as views of the U.S. Department of Defense, the U.S. Department of the Army, or the U.S. Department of Justice.