A facile cost-efficient fast and reliable two-stage strategy for the high-throughput identification of samples belonging to mtDNA haplogroup K and its subhaplogroups K1 and K2

Harald Niederstätter, Christiane Bauer, Liane Fendt, Walther Parson

Institute of Legal Medicine, Innsbruck Medical University, Austria

Summary We developed for a study focusing on the mitochondrial haplogroup K (hg K) and its subhaplogroups K1 and K2 a two-staged reliable low-cost assay for high throughput identification of samples attributable to these haplogroups in a large Austrian population sample. For this purpose homogeneous allele-specific PCR (the amplification refractory mutation system, ARMS) was used. With hg K respectively non-hg K samples the ARMS specifically amplified either the derived 9055A allele (hg K samples) or the ancestral 9055G allele (non-hg K samples) in the presence of the dsDNA binding fluorescent dye Eva Green. Further subdividing the samples attributed to hg K was performed using an competitive duplex ARMS assay simultaneously interrogating the SNP alleles 1189C (hg K1) and 9716C (hg K2). To differentiate non-hg(K1,K2) or potential K* samples from PCR failure, a short sequence stretch around nucleotide position (ntp) 9055 was co-amplified with the two ARMS products. For both assays, product detection and allele calling were based on the amplicon characteristic melting-point temperatures (T_m) obtained by on-line post-PCR fluorescent dissociation curve analysis (DCA). Both ARMS-DCA assays worked over a broad range of initial DNA concentrations, and enabled the successful analysis of a large population sample without using robotic lab-equipment. We conclude that homogeneous fluorescent ARMS-DCA in general represents a very attractive methodological approach for large-volume studies when the number of SNP markers is limited. It might be also useful for the fast exclusion of the vast majority of specimens in mass-screenings if the haplogroup affiliation of the offender is known.

Materials and methods The ARMS consists solely of PCR amplification and can be used for the detection of any known base substitution and small insertions/deletions. In general, ARMS utilizes two (three) primers, one being non-allele specific and the other one (two) with a 3' terminal base specific for the interrogated allelic variant (Figs. 1, 2). The specificity of the reaction depends on the inability of *Taq* polymerase to extend 3'terminally mismatched oligonucleotides but not on the length of the primers, which makes primer design straightforward. However, *Taq* polymerase extends primers despite a 3'terminal mismatch, albeit at a much lower rate. In order to completely prevent *Taq* polymerase from extending 3'terminally mismatched primers, a deliberate mismatch was introduced at the 3' antepenultimate nucleotide position of the allele-specific primers (Figs.1, 2). Thus, the allele-specific primers displayed a single mismatch with their target allele, two mismatches with the alternative allelic variant and zero mismatches with their PCR products, ensuring both, specificity as well as sensitivity. For the set up of the reaction mixture and the thermal cycler protocols refer to Tables 1 and 2 and Figure 3. For the typing of the population sample - comprising DNA samples extracted from peripheral blood donated by 3,680 healthy individuals from Tyrol (Austria) after informed consent - ARMS was performed off-line in standard thermal cyclers. Following amplification, the PCR products were scored according to their characteristic T_m values, which were derived from on-line monitoring of the EvaGreen or SYBR Green I fluorescence during a temperature ramp from 60°C to 90°C (Figs. 3, 4, 5) using a real-time PCR cycler (Applied Biosystems 7700 or 7500Fast).

Population study results The stepwise mtSNP typing approach for the identification of hg K samples and their subdivision into hgs K1, K2 was applied to a large sample set comprising 3,680 DNA extracts from blood donors living in Tyrol (Austria, Fig. 6). 351 of the 415 samples initially attributed to hg K belonged to hg K1 and 59 could be scored K2. The remaining 5 specimen were typed to be non(K1,K2). Sequencing analysis of these samples revealed in all instances a 9055A homoplasy, causing erroneous hg K assignments (Table 3). Additionally, 5 samples yielded ambiguous hg K vs. non-K results due to sequence variation in the 9055G/A forward primer binding site (Table 4). None of them belonged to hg K. Thus, in total the Tyrolean population sample comprised of 3,270 non-hg K and 410 hg K individuals (11.14%). All of the hg K1 and K2 subtyping results were confirmed by sequencing analysis.

Fig. 1 hg K/non-K amplicons and primer sequences

hg K vs. non-hg K

```
9055G F: gcggcCTCATGCACCTAATTGGAAcCG
9055A F:      ACTCATGCACCTAATTGGAAtCA
rCRS:         ACTCATGCACCTAATTGGAAgCGccaCCCTAGCAATATCAACCATTAACCTTCCC
9055 R:                                  GGGAAGGTTAATGGTTGATATTGCTAGGG
```

Table 1 hg K/non-K ARMS cocktail

non-acetylated BSA	250 ng/µL
9055G F	150 nM
9055A F	150 nM
9055 R	150 nM
AmpliTaq Gold (Applied Biosystems, AB)	0.1 U/µL
EvaGreen qPCR Basic Mix HS (Biotum)	1x
assay volume	10 µL

Fig. 4 hg K/non-K 1st derivative amplicon melting curves

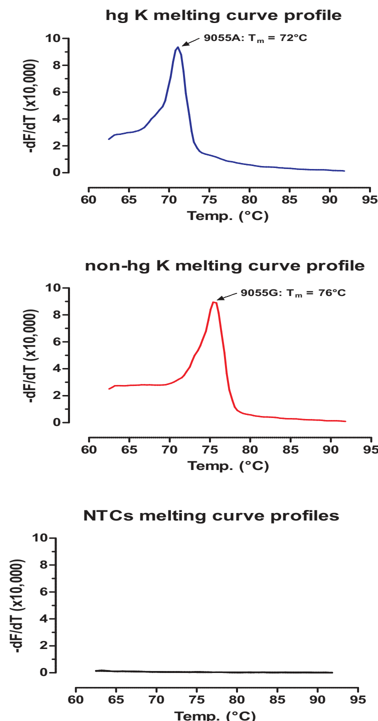


Fig. 3 Cycler profile and fluorescence signal plots

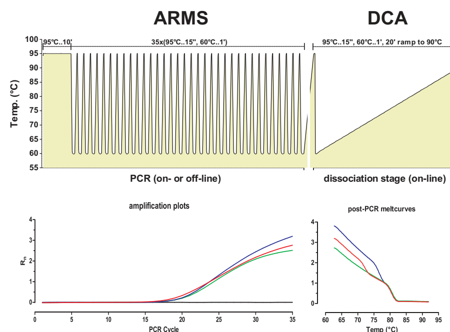


Fig. 6 population sample typing results

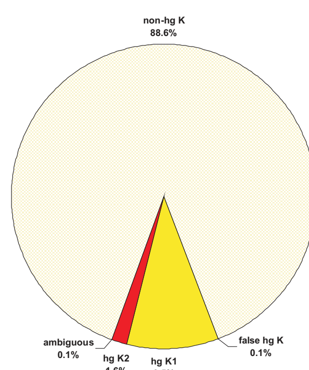


Table 3 false hg K assignments

hg K vs. non-K typing result	seq. variation forward primer	seq. variation reverse primer	mtDNA hg	n (%)
hg K	9055A	-	J1c1	1 (0.03)
hg K	9055A	-	H	1 (0.03)
hg K	9055A	-	U5a1	1 (0.03)
hg K	9055A	-	HV0	2 (0.05)

Table 4 ambiguous hg K vs. non-K typing results

hg K vs. non-K typing result	seq. variation forward primer	seq. variation reverse primer	mtDNA hg	n (%)
9055G/A double peak	9055R	-	H1a	1 (0.03)
9055G/A double peak	9053A, 9055R	-	HV	1 (0.03)
9055G/A double peak	9052G	-	J1c	2 (0.05)
9055G/A double peak	9052R	-	J1c	1 (0.03)

Fig. 2 hg K1, K2, IPC amplicons and primer sequences

hg K1

```
1189C F: ACCTGGCGGTGCTTCAgAC
rCRS:    ACCTGGCGGTGCTTCAtATccctctagaggagcctgttctgtaatcGATAAACCCCGATCAACCTC
1189 R:                                                  CTCCAACTAGCCCCAAATAG
```

hg K2

```
9716 F:  CTAATAGAAAACAACCGAAACCAAA
rCRS:    CTAATAGAAAACAACCGAAACCAAAtaattcaagcactgcttattacaattttactgggTCtCTATTTTACCCTCCTACAAGC
9716C R:                                                            GGtGATAAAATGGGAGGATGTTCG
```

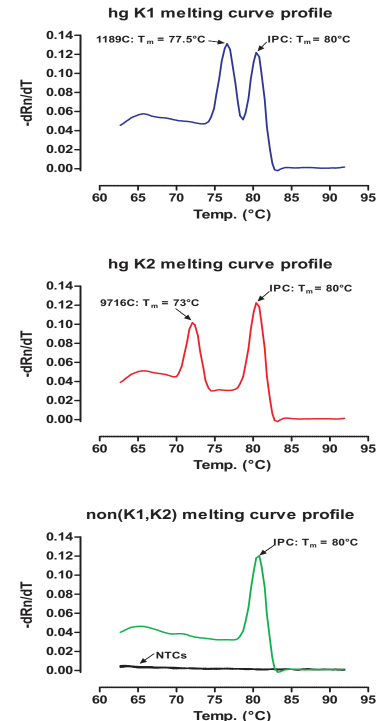
IPC

```
5973 F:  GCATGAGCTGGAGTCCTAGGCACAG
rCRS:    GCATGAGCTGGAGTCCTAGGCACAGctctaagcctccttattcgagccgagctgggccagccaGGCAACCTTCTAGGTAACGACCACATCTACA
6066 R:                                                                  CCGTTGGAAGATCCATTGCTGGTGTAGATGT
```

Table 2 hg K1, K2, IPC ARMS cocktail

non-acetylated BSA	250 ng/µL
1198C F	250 nM
1189 R	250 nM
9716 F	375 nM
9716C R	375 nM
5973 F	250 nM
6066 R	250 nM
Power SYBR Green PCR Master Mix (AB)	1x
assay volume	10 µL

Fig. 5 hg K1, K2, IPC 1st derivative amplicon melting curves



Acknowledgements: Part of this work was financially supported by a grant within the framework provided by the "Alpine Research" programme of the Austrian Academy of Sciences.