Int J Legal Med (2003) 117 : 291–298
DOI 10.1007/s00414-003-0395-2

ORIGINAL ARTICLE

Anita Brandstätter · Thomas J. Parsons · Walther Parson

Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups

Received: 20 May 2003 / Accepted: 23 June 2003 / Published online: 21 August 2003

Abstract This work presents a selection of 16 SNPs from the coding region of the human mitochondrial DNA. The selected markers are used for the assignment of individuals to one of the nine major European Caucasian mitochondrial haplogroups. The selected SNPs are targeted in two multiplex systems, via the application of the SNaPshot kit, a multiplex method based on the dideoxy single-base extension of unlabeled oligonucleotide primers. The method is conceived as a rapid screening technique prior to sequencing analysis, in order to eliminate multiple suspects from an inquiry or to discriminate between stains in a high volume casework example. Moreover, the ability to assign an unknown sample to an mtDNA type of known Caucasian origin could be of probative value in some investigations. A database of 277 Austrian Caucasians has been constructed, and the probability of a chance match between two unrelated individuals is calculated as 11.4%. This novel multiplex PCR amplification and typing system for mtDNA coding region SNPs promises to be a convenient and informative new DNA profiling system in the forensic field.

Keywords Caucasian mtDNA haplogroups · Phylogenetic SNPs · Multiplex SNP typing · Primer extension method · Forensic science

Introduction

Analysis of single nucleotide polymorphisms (SNPs) – the most common form of variation in the human genome – has become a popular strategy for discovering genes involved in complex diseases [1, 2, 3] and its promising application in forensic human identification has been recognized [4, 5, 6]. Whereas techniques for determination of the molecular bases of disease states normally involve samples with large amounts of high quality DNA, forensic analysts are often confronted with human identification cases where the amount and quality of nuclear DNA are greatly reduced. In many cases, the amplification of conventional STRs fails or gives inconclusive results. Typing of the mitochondrial DNA control region (or their hypervariable parts) is then the method of choice because the high copy number of mtDNA ensures a better PCR amplification success rate over nuclear loci. The consequence of its mode of inheritance (maternal transmission and lack of recombination) is a lower power of discrimination, the most common HV1/HV2 haplotype is found in approximately 7% of the Caucasian population. In order to compensate for the weak discrimination of common HV1/HV2 haplotypes in routine casework, further sequence analysis of additional mtDNA control region fragments has been performed [7]. However, this technique requires more DNA extract and is time consuming. An alternative approach for further differentiation of identical mtDNA haplotypes is the analysis of single nucleotide polymorphisms (SNPs) which are located in the coding region of the mtDNA [8]. We based our selection of markers on the ability to ascribe samples to one of the major west European Caucasian mtDNA haplogroups. By screening these SNPs prior to sequencing analysis of the hypervariable regions we would be able to rapidly differentiate between stains or hairs in high volume casework or to eliminate multiple suspects from an inquiry. For multiplexing of the selected SNPs, the SNaPshot primer extension technique was used as a rapid screening method.

SNaPshot (AB, Applied Biosystems, Foster City, CA) is a commercially available mini-sequencing method that relies on the single-base extension of a primer immediately adjacent to the SNP using fluorescently labelled ddNTPs. The fluorescently labeled extension products can then be separated and visualized by electrophoresis and fluores-

A. Brandstätter · 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

T. J. Parsons
Armed Forces DNA Identification Laboratory,
Armed Forces Institute of Pathology,
Building 101, 1413 Research Boulevard,
Rockville, MD 20850, USA

cence detection. As four different dyes can be detected simultaneously, any base position can be assessed for the presence of an A, C, G, or T base. MtDNA mini-sequencing was found to provide a reliable and reproducible technique for forensic applications [9], and the capability of simultaneously identifying numerous mtDNA substitution polymorphisms within a single reaction gives SNaPshot high forensic potential [10]. The 2-stage strategy of screening samples using SNaPshot multiplex assays, followed by sequencing a small number of matching samples would permit mtDNA analysis to be used in a much wider range of forensic cases than would be possible using sequencing alone.

Materials and methods

DNA samples and extraction

Blood samples from 277 unrelated Austrian Caucasians were Chelex extracted as described before [11].

Primer design and evaluation

All primers were designed using the program Primer Express (Version 1.5, AB), secondary structures were calculated using the web-run program mfold 3.0 [12] (http://www.bioinfo.rpi.edu/applications/mfold/old/dna/), and possible self/cross-extensions were checked using the program MeltCalc [13]. For SNP interrogation using primer extension, the 3'-end of a primer has to hybridize one base to the 5'-side of the polymorphic position. The electrophoretic mobility of extension products can be adjusted by varying the length of the extension primer using 5' non-homologous tails. Varying the binding site between the sense and antisense strands was used to divide peaks into the four possible channels. Mini-sequencing primers were tested experimentally in a "self-extension reaction" using the Primer Focus kit (AB), where no PCR template was added to the primer extension mixture.

PCR multiplex amplification and mini-sequencing

High-throughput genotyping was established by dispensing aliquots of the DNA extracts in 96-well plates, and 4 negative control samples were assayed per plate. For amplification, to each well of a MicroAmp Optical 96-well reaction plate (AB), 19 µl of PCR master mix containing 1.0 unit of Advantage 2 (Clontech, BD Biosciences, Palo Alto and Mountain View, CA), 1.0 unit of reaction buffer (Clontech) and 200 µM each dNTP (AB) were dispensed. Amplification primers and their concentrations are specified in Table 1. Then 1 µl DNA extract (containing 0.5–25 pg of genomic DNA) was put into the PCR master mix with a multichannel pipettor. The amplification reaction was conducted on an iCycler (Biorad, Hercules, CA) in two 8-plex PCR reactions. Multiplex I comprised the SNP target sites G709A, G1719A, G3010A, C7028T, A11251G, G12372A, T14798C, and C15904T. Multiplex II comprised the SNP target sites A1811G, T6365C, T6776C, G8251A, G8697A, G9055A, G13708A, and C14766T. The first letter denotes the L-strand nucleotide in the rCRS (revised Cambridge reference sequence; [14]); the second letter denotes the non-rCRS allele. After an initial denaturation at 95°C for 2 min, the reactions were put through 28–32 cycles of denaturation at 95°C for 30 s, annealing at 66°C for 30 s, and extension at 72°C for 30 s. PCR primers and unincorporated dNTPs were removed by adding ExoSAP-IT (USB, Cleveland, OH) according to the manufacturer's protocol. The extension primers (Table 2) were premixed for each multiplex master mix, giving a final concentration of 0.2 µM for each primer. The reactions comprised 2.5 µl of the SNaPshot mul-

Table 1 Primer sequences and concentrations for PCR amplification

SNP	Forward primer	Reverse primer	Final primer concentration in reaction mix
G709A	5'-*TAGCTCTTAGTAAGATTACACATGCAAGCATCCC*-3'	5'-*TGCTTGTCCCTTTTGATCGTGGTGA*-3'	100 nM
G1719A	5'-*CTAAACCTAGCCCCAAACCCACTCCAC*-3'	5'-*CAGGTTTCAATTTCTATCGCCTATACTTTATTTGGGT*-3'	50 nM
A1811G	5'-*TGAAACCTGGCGCAATAGATATAGTACCGC*-3'	5'-*CATTATGCAGAAGGTATAGGGGTTAGTCCTTGC*-3'	150 nM
G3010A	5'- *CGACCTCGATGTTGGATCAGGACA*-3'	5'- *CGAACCTTTAATAGCGGCTGCACCAT*-3'	75 nM
T6365C	5'-*CCTGGAGCCTCCGTAGACCTAACCATCT*-3'	5'-*TTGTTGTGATGAAATTGATGGCCCCTAA*-3'	125 nM
T6776C	5'-*GTCTGAGCTATGATATCAATTGGCTTCCTAGGGT*-3'	5'-*CGGAGGTGAAATATGCTCGTGTGTCTACG*-3'	75 nM
C7028T	5'-*GCAAACTCATCACTAGACATCGTACTACACGACA*-3'	5'-*TGAAGCCTCCTATGATGGCAAATACAGC*-3'	175 nM
G8251A	5'-*CCACAGTTTCATGCCCATCGTCCTAG*-3'	5'-*CTCTAGAGGGGGTAGAGGGGGTGCTATAGG*-3'	125 nM
G8697A	5'-*AACAACCGACTAATCACCACCCAACAATG*-3'	5'-*GAGATCAGGTTCGTCCTTTAGTGTTGTGTATGG*-3'	75 nM
G9055A	5'-*CGTACGCCTAACCGCTAACATTACTGCAG*-3'	5'-*AGTGTAGAGGGAAGGTTAATGGTTGATATTGCTA*-3'	125 nM
A11251G	5'-*CAGCCAGAAGCGCCTGAACGCA*-3'	5'-*GAATGTTTAGTGAGCCTAGGGTGTTGTGAGTGT*-3'	65 nM
G12372A	5'-*TGCAACTCCAAATAAAAGTAATAACCATGCACA*-3'	5'-*GGGTTAACGAGGGTGGTAAGGATGGG*-3'	50 nM
G13708A	5'-*ACCCTTACTAACATTAACGAAAATAACCCCACCCT*-3'	5'-*AAATCCTGCGAATAGGCTTCCGGCT*-3'	50 nM
C14766T	5'-*ATCGTTGTATTTCAACTACAAGAACACCAATGACC*-3'	5'-*GTGGGGAGGTCGATGAGTGAGTGGTTA*-3'	100 nM
T14798C	5'-*CCCCAATACGCAAAATTAACCCCCTAA*-3'	5'-*CGAAGTTTCATCATGCGGAGATGTTGG*-3'	75 nM
C15904T	5'-*CCCTAATTGAAAACAAAATACTCAAATGGGCC*-3'	5'-*AGGTTTTCATCTCCGGTTTACAAGACTGGTG*-3'	75 nM

Table 2 Extension primers

SNP	Tail and primer sequence	Incorporated ddNTP
G709A	gac(gact)$_5$*TAGAGGGTGAACTCACTGGAA*	C~T
G1719A	gac(gact)$_2$*CCACCTTACTACCAGACAACCTTA*	G~A
A1811G	(gact)$_4$*GCAAGGGAAAGATGAAAAATTATA*	A~G
G3010A	ga(gact)$_8$*TTAATAGCGGCTGCACCAT*	C~T
T6365C	(gact)$_2$*CATCTTCTCCTTACACCTAGCAGG*	T~C
T6776C	ga(gact)$_4$*GTGTGTCTACGTCTATTCCTACTGTAAATAT*	A~G
C7028T	gac(gact)$_{10}$*CTACACGACACGTACTACGTTGTAGC*	C~T
G8251A	(gact)$_6$*GGGTGCTATAGGGTAAATACGGG*	C~T
G8697A	gac(gact)$_8$*CGTCCTTTAGTGTTGTGTATGGTTAT*	C~T
G9055A	gac(gact)$_1$*ACTCATGCACCTAATTGGAAGC*	G~A
A11251G	(gact)$_1$*CTTCCCCTACTCATCGCACT*	A~G
G12372A	(gact)$_5$*CACACTACTATAACCACCCTAACCCT*	G~A
G13708A	*TAAACCCCATTAAACGCCTG*	G~A
C14766T	ga*AATGACCCCAATACGCAAAA*	C~T
T14798C	*CCCCTAATAAAATTAATTAACCACTCA*	T~C
C15904T	g(gact)$_3$*CCGGTTTACAAGACTGGTGTATTA*	G~A

Primer tails are printed in lower case letters, the matching primer sequences are written in italic upper case letters.

Table 3 Frequencies of previously established haplogroup-specific polymorphisms

Nucleotide substitution	H (226) 0.52[a]	I (14) 0.032[a]	J (33) 0.076[a]	K (47) 0.108[a]	T (46) 0.106[a]	U (42) 0.097[a]	V (8) 0.018[a]	W (8) 0.018[a]	X (11) 0.025[a]	Relative frequency
G709A	*0.013*	*0*	*0*	*0.191*	*1*	*0.024*	*0*	*1*	*0*	0.154[b]
G1719A	*0.009*	*1*	*0.030*	*0*	*0.022*	*0*	*0*	*0*	*1*	0.067[b]
A1811G	*0.004*	*0*	*0*	*0.979*	*0*	*0.357*	*0*	*0*	*0*	0.143[b]
G3010A	*0.323*	*0*	*0.818*	*0*	*0*	*0.071*	*0*	*0*	*0*	0.237[b]
T6365C	*0.044*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	0.023[b]
T6776C	*0.124*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	0.064[b]
C7028T	*0.053*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	0.508[b]
G8251A	*0*	*0.929*	*0*	*0*	*0.022*	*0*	*0*	*1*	*0*	0.050[b]
G8697A	*0*	*0.071*	*0*	*0.021*	*1*	*0*	*0*	*0*	*0*	0.111[b]
G9055A	*0.004*	*0*	*0*	*1*	*0*	*0*	*0*	*0*	*0*	0.110[b]
A11251G	*0*	*0*	*1*	*0*	*1*	*0*	*0*	*0*	*0*	0.182[b]
G12372A	*0*	*0*	*0*	*1*	*0*	*1*	*0*	*0*	*0*	0.205[b]
G13708A	*0.018*	*0*	*1*	*0*	*0*	*0.024*	*0*	*0*	*0.455*	0.099[b]
C14766T	*0.008*	*0.785*	*1*	*0.978*	*1*	*1*	*0*	*1*	*1*	0.457[b]
T14798C	*0*	*0*	*0.727*	*1*	*0*	*0*	*0*	*0*	*0*	0.163[b]
C15904T	*0*	*0*	*0*	*0*	*0*	*0*	*1*	*0*	*0*	0.018[b]

Italics indicate the observed frequency of the polymorphism inside the particular haplogroup.
[a]Observed incidence of the particular haplogroup in the European population.
[b]Observed frequency of the polymorphism inside the European population.
All nucleotide substitutions are indicated relative to the rCRS and are designated by two letters bracketing a number: the first letter denotes the L-strand nucleotide in the rCRS (revised Cambridge reference sequence; [14]), and the second letter denotes the non-rCRS allele. Data were compiled and summarized from [17].

tiplex ready reaction mix, 2.0 µl of PCR product, 1.0 µl of pooled extension primers, and distilled water up to 10 µl. Thermal cycling and post-extension treatment were conducted following the manufacturers protocol. Unincorporated ddNTPs were removed with SAP (shrimp alkaline phosphatase, USB).

Electrophoresis and data analysis

For preparation of samples for electrophoresis, 2 µl of SNaPshot product was mixed with 12 µl of Hi-Di formamide, containing 0.58 µl of GeneScan-120 LIZ size standard per 100 µl of Hi-Di formamide. Finally 2 µl of the diluted SNaPshot product was mixed with 8 µl of the GeneScan-120 LIZ size standard-Hi-Di formamide mixture. Samples were denatured at 95°C for 5 min. The electrophoresis was run on an ABI 3100 Genetic Analyzer using 3100-POP-4 polymer (AB). Files were analyzed using GeneScan Analysis Software version 3.1 and GeneScan-120 LIZ size standard analysis parameter files. Results were evaluated with Genotyper version 2.5 (all AB).

Random matching probability

The probability of a chance match (M) between two unrelated individuals was calculated as: $P(M)=\frac{1}{n^2}\sum_{i=1}^{m} x_i^{\,2}$ [15], where n is the number of individuals in the database, m is the number of haplotypes in the database and x_i is the number of times the i^{th} haplotype was seen in the database.

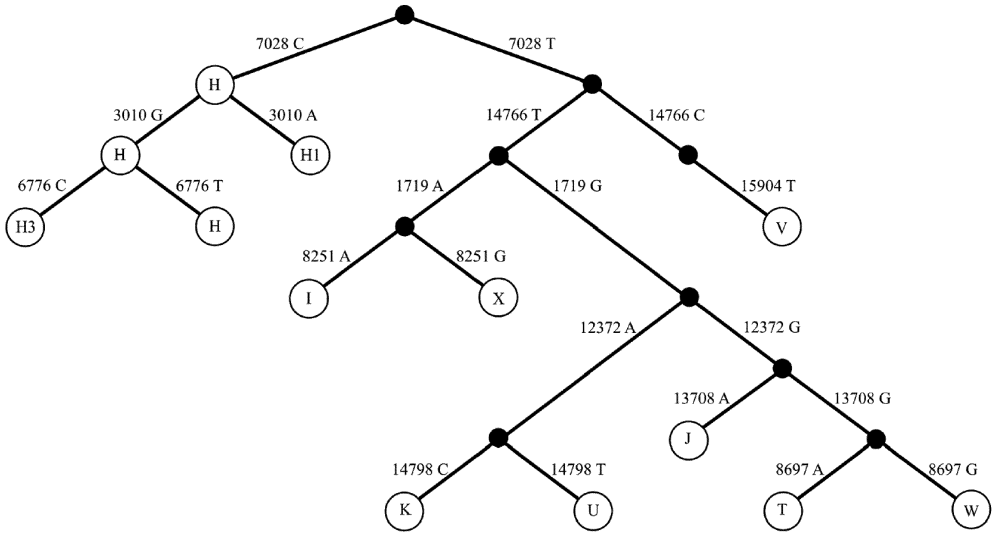

Fig. 1 Bifurcating decision tree for haplogroup assignment

Haplogroup assignment

Our genotyping strategy is based on the known phylogenetic network of the mitochondrial genome. From publications on the phylogenetic history and geographical origin of human populations [16, 17, 18, 19], we chose 16 candidate SNP sites, which in combination are useful for the characterization of a haplogroup and further have a reasonable frequency distribution in the populations with respect to all nine major European haplogroups (H, I, J, K, T, U, V, W, X) (Table 3). Individuals can be assigned to haplogroups by following the branches of a bifurcating tree (Fig. 1).

Results and discussion

The evolution of the human mitochondrial genome is characterized by the emergence of distinct lineages or haplogroups which are clusters of evolutionary closely related mtDNA lineages. Nine European, seven Asian (including Native American), and three African mtDNA haplogroups have been identified previously on the basis of the presence or absence of restriction enzyme recognition sites [16, 20, 21], by means of nucleotide sequences within the control region [22, 23], and recently, as a result of complete mtDNA sequences [17, 18].

Although forensic applications of human mtDNA variation outside the control region have been rare until recently [8, 24], a growing number of potential SNP assay sites are now available from human evolution studies [17, 25, 26]. Based on information available in the literature, one can identify binary markers that define a well-supported and stable phylogenetic tree, which allows a simple screening strategy to be devised. We designed our typing method with four features in mind. Firstly, we needed to type markers which define all nine major west European Caucasian haplogroups. Secondly, we needed additional markers which further discriminate within the haplogroups to raise the discrimination power. Thirdly, the mutations at the sites selected should not have bio-functional consequences, i.e. silent mutations on third codon positions were preferred. Fourthly, amplification was designed for very short DNA fragments (50~100 base pairs) to raise the success rate for degraded DNA.

SNP genotyping

All 277 samples were successfully typed with the two multiplexes, giving clear results for all 16 markers typed (Fig. 2). The results of the SNaPshot reactions were confirmed by direct sequencing of the PCR products using the control DNA 9947A (Promega Corporation, Madison, WI). Macro-assisted evaluation of the raw data, with macros written for Genotyper software (AB) and Excel (Microsoft, Redmont, WA) enabled a very fast and correct assessment of the individual profiles (the Genotyper macro will be sent to interested researchers on request). Within the multiplex II, two extraneous peaks were observed in nearly all cases: in many cases there was a peak sized ~20 bases near the 13708A peak in the green channel,, and in the red channel there was always an additional artefactual peak ~5 bases 5' to the 6365T peak. Causes for these extraneous peaks are unknown, but may include primer hairpin extension and primer dimer extension. However, these artefacts did not affect the SNP interrogation process, as the Genotyper-macro disabled a labelling of peaks of that size and colour, and no other peak was expected to appear there.

Although heteroplasmy was not specifically tested, we observed the coexistence of two populations of mtDNA molecules with different nucleotide composition in two individuals at one position each (Table 4 lines 3 and 7).

The length of amplified DNA segments varied between 70 and 100 base pairs, enabling short DNA fragments to be amplified. The method proved to be very sensitive: the major part of the samples (with average amounts of genomic DNA of 25 pg per assay) was typed by use of 28 amplification cycles; and even with very little amounts of genomic DNA (1 pg per assay) a full profile was obtained by increasing the number of PCR amplification cycles to 32. The sensitivity of the assay was tested in a preliminary study

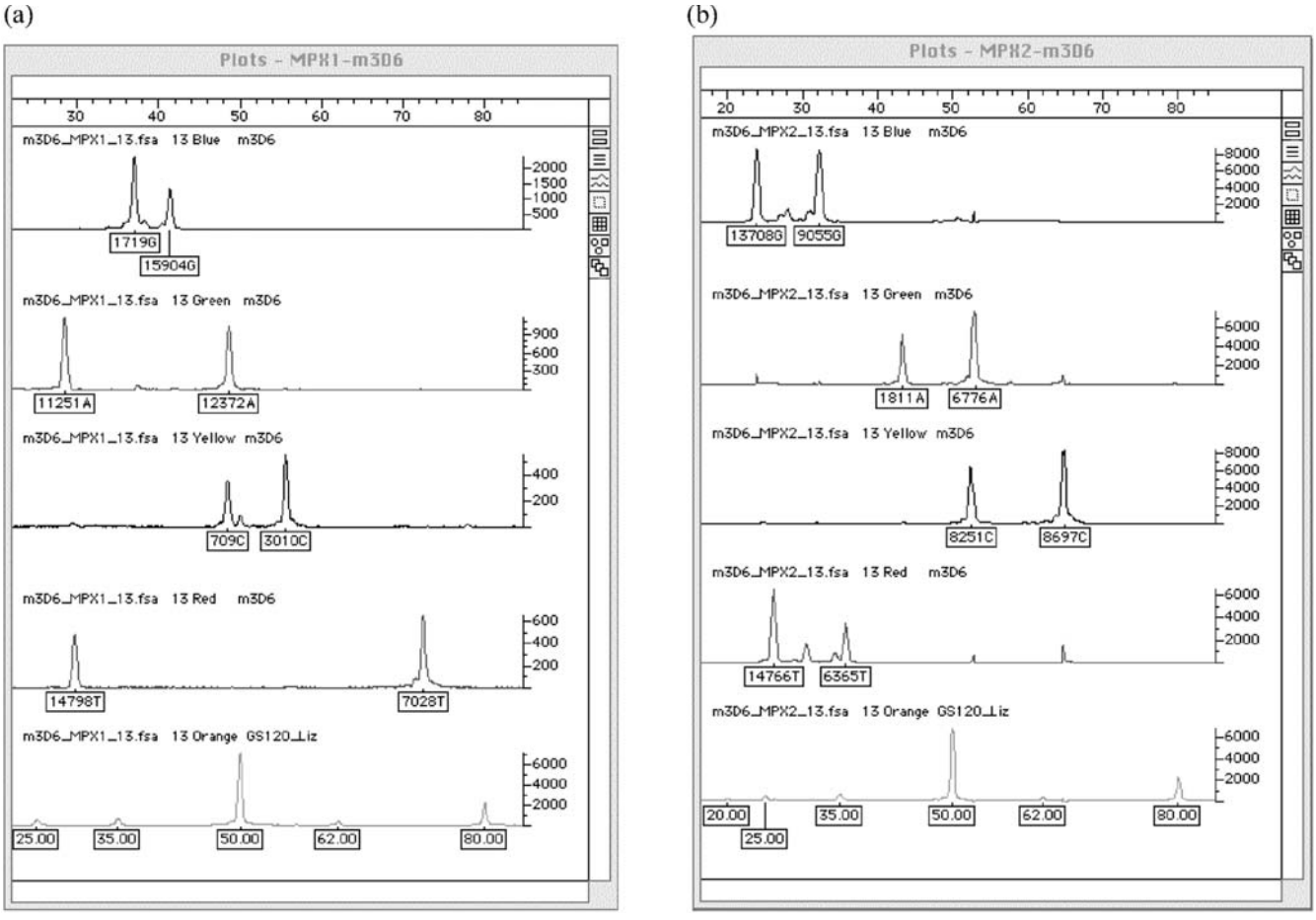

Fig. 2a, b Example of results for the two SNaPshot multiplexes from the same sample (assigned to haplogroup U). **a** Multiplex I, **b** Multiplex II

(data not shown) by amplifying a dilution series (0.5–100 pg DNA) of the control DNA 9947A (Promega; 10 ng/µl).

Haplogroup assignment

All haplotypes found in this study are summarized in Table 4. All nine major West-European Caucasian haplogroups (and two sub-haplogroups) were observed in our sample, and only one individual did not fit into these predefined haplogroups and was classified as "others". This individual most likely belongs to a rare west European haplogroup (e.g., R or Z) [25] or to the historical admixture known to exist in the European population [27]. The phylogenetic tree is bifurcated by the presence or absence of the C7028T nucleotide substitution (Fig. 1). European mtDNAs lacking the C7028T site fall into haplogroup H, those retaining the site fall into the haplogroups: I, J, K, T, U, V, W, and X. Haplogroup H encompassed 45.0% of the Austrian Caucasians, and a further subdivision could be established by the presence of the G3010A site (*H1*; 14.5%) or the presence of the T6776C polymorphism (*H3*; 2.3%). Haplogroup I is defined by the presence of nucleotide substitutions G1719A, G8251A and C14766T and accounts for 1.4% of Austrian mtDNA sequences. Haplogroup J is defined by the presence of the polymorphisms A11251G, G13708A and C14766T and accounts for 9.0% of the studied mtDNAs. Haplogroup K is delineated by transitions A1811G, G12372A, C14766T, and T14798C and is found in 8.3% of the samples. Haplogroup T is defined by the nucleotide substitutions G709A, G8697A, A11251G and C14766T and is found in about 13.2% of Austrian Caucasians. Haplogroup U is defined by the presence of the G12372A and the C14766T polymorphisms and represents 18.8% of Austrian mtDNAs. Haplogroup V is identified by the presence of the C15904T site and is found in about 1.4% of Austrian mtDNAs. Haplogroup W is found in 1.4% of Austrians and is delineated by three characteristic polymorphisms: G709A, G8251A, and C14766T. Haplogroup X (1.4% of Austrian Caucasians) is characterized by the two polymorphic positions 1719 G→A and C14766T and can be distinguished from haplogroup I by the lack of the G8251A polymorphism.

These proportions are in agreement with the findings from surveys of European populations [17, 20, 21, 25] (Table 5).

Phylogenetic analyses of the mtDNA control region sequences (sequences provided by H. Niederstätter, Institute of Legal Medicine, University of Innsbruck) yielded the same results as the haplogroup-identifications via the 16 coding region SNPs; the CR sequences were assigned

Table 4 mtDNA coding region haplotypes found in 277 Austrian Caucasians

Haplogroup	Haplotype	Number
H	7028C	69
H	No difference	6
H	7028C – 8697R	1
H	7028C – 9055A	1
H	7028C – 709A	1
H1	3010A – 7028C	34
H1	709A – 3010A – 7028Y	2
H1	3010A – 6365C – 7028C	2
H1	3010A – 7028C – 11251G	1
H1	3010A – 7028C – 8251A	1
H3	6776C – 7028C	4
H3	3010A – 6776C – 7028C	1
H3	6776C – 7028C – 9055A	1
I	1719A – 14766T	2
I	1719A – 8251A – 14766T	2
J	3010A – 11251G – 13708A – 14766T – 14798C	17
J	11251G – 13708A – 14766T	4
J	3010A – 11251G – 13708A – 14766T	3
J	709A – 3010A – 11251G – 13708A – 14766T – 14798C	1
K	1811G – 9055A – 12372A – 14766T – 14798C	14
K	709A – 1811G – 9055A – 12372A – 14766T – 14798C	5
K	1811G – 8251A – 9055A – 12372A – 14766T – 14798C	2
K	709A – 1811G – 9055A – 12372A – 13708A – 14766T – 14798C	1
K	709A – 1811G – 12372A – 14766T – 14798C	1
T	709A – 8697A – 11251G – 14766T	31
T	709A – 3010A – 8697A – 11251G – 14766T	3
T	709A – 8697A – 11251G – 14766T – 14798C	1
T	709A – 8251A – 8697A – 11251G – 14766T	1
U	12372A – 14766T	26
U	1811G – 12372A – 14766T	23
U	8251A – 12372A – 14766T	2
U	1811G – 3010A – 12372A – 14766T	1
V	15904T	3
V	1719A – 15904T	1
W	709A – 8251A – 14766T	4
X	1719A – 14766T	3
X	1719A – 13708A – 14766T	1
Others	14766T	1

to haplogroups by use of the diagnostic polymorphisms that have previously been reported [16, 28] (data not shown), and thus confirmed the assignment of the samples to the haplogroups. However, haplogroups I and X could only be distinguished by the presence of the G8251A site (haplogroup I) or by the use of control region sequences. In this study, two individuals were clearly assigned to haplogroup I by characteristic control region sites, but lacked the G8251A site. This situation was expected to happen, as the frequency of this polymorphism in haplogroup I was calculated as 92% (Table 3). But still, in many cases it may be sufficient to know that the defendant derives from the super-haplogroup IX, as this cluster accounts only for 2.8% of the Austrian Caucasian population.

Random matching probability

The most common haplotype found was a profile matching the rCRS, thus assigned to haplogroup H (24.5%). Overall, 37 different lineages were found in 277 individuals, and 16 of them appeared only once in the dataset. The probability of a random match between two unrelated individuals was calculated as 11.4%.

Mitochondrial polymorphisms and disease

Many mitochondrial polymorphisms are known to be associated with diseases, and certain European mtDNA haplogroups are prone to a genetic susceptibility for various disorders [29, 30]. It is also widely anticipated that SNPs will play a major role in discovering genes involved in complex diseases such as Alzheimer's disease, obesity and

Table 5 Haplogroup frequencies in previously published studies

Haplo-group	Austrians (this study) (*N*=277) %	Finns [25] (*N*=480) %	Scandinavians [20] (*N*=86) %	US Americans [21] (*N*=259) %	US Americans [17] (*N*=435) %
H	45	39	41.7	40.5	52
I	1.4	3	1.2	6.7	3.2
J	9.1	5	9.3	11.3	7.6
K	8.3	3	8.1	9.1	10.8
T	13	3	12.8	15.2	10.6
U	18.8	28	17.3	14.7	9.7
V	1.4	6	3.6	4.8	1.8
W	1.4	10	2.3		1.8
X	1.4	1	2.3	6.9	2.5

diabetes. Reflecting these expectations, the haplogroup defining SNPs for this study were selected carefully in terms of having poor medical-genetic information, i.e. predominantly third codon positions or positions with conservative amino acid changes were considered for genotyping. As a matter of fact, a recent study [30] reported a correlation of polymorphisms 9055 G → A and 13708 G → A with a mild protective effect against Parkinson's disease among women and elderly people, respectively. Although we might never be protected against new findings of mitochondrial dysfunctions, we are optimistic that this multiplex system provides a good basis for haplogroup assignment, without unintentionally becoming a tool for disease screening.

Forensic assessment

One of the discussions in the forensic community is the number of STRs that are needed to be typed in order to achieve the maximum exclusion probability in practical terms. This issue is even more significant when mtDNA markers are in question, due to the non-recombining nature of the genome. It is accepted that the addition of an informative mtDNA marker to the classical HV1/HV2 sequence may increase the power of discrimination up to a certain point where it will not be possible to discriminate more haplotypes and any extra information becomes redundant. Therefore our strategy to tackle the problem was to select a set of markers which meet certain conditions: to discriminate between major European mtDNA lineages, to have a substantial power of exclusion with a minimal set of SNPs that may be targeted in two multiplex reactions, and to be sensitive for degraded DNA or for low amounts of intact DNA. The screening procedure presented here proved to be practical for unambiguously assigning haplogroup origin to unknown DNA samples, and the method itself was shown to be very sensitive and suitable to type samples containing at least 1 pg genomic DNA. The entire procedure may be completed in 1 day and integrates easily with existing laboratory equipment. In addition, the assay is well suited for use on highly degraded samples, as the amplicon lengths do not exceed 100 bp.

For comparison to an alternative approach for rapid mtDNA screening, we calculated the power of discrimination for the 277 unrelated Austrian Caucasians with the SSO binding sites as described in [31]. The SSO system targets 15 mtDNA control region markers. The calculated POD was 91.7%, compared to a POD of our SNP panel of 88.6%. The slightly lower discrimination power from our panel is offset by the fact that the SNaPshot panels provide information that augments the information in the control region. Therefore subsequent HV1/HV2-sequencing will not lead to redundant results, as when CR sites are targeted. Apart from the ability to eliminate non-probative DNA samples from further investigations, the assignment of a sample to a certain haplogroup gives the investigator the possibility to target specific control region fragments, which show a higher variability for certain haplogroups. Haplogroup U, for example, can be divided into its sub-haplogroups U2 and U4 by sequencing HV3 [32] of the CR (U2 is characterized by CR polymorphisms 340T and 508G, U4 can be distinguished by the presence of the CR site 499A). Another example is haplogroup T: this haplogroup shows its highest CR variability in HV1, whereas HV2 shows mainly the "common" polymorphisms 73G and 263G. The information of the haplogroup affiliation of an unknown sample, in combination with the knowledge of sequence variability in the entire control region enables the investigator to selectively target specific control region sites, and this strategy has the potential to reduce reagent expense and sample consumption. This novel, multiplex PCR amplification and typing system for 16 mtDNA coding region SNPs promises to be a convenient and informative new DNA profiling system in the forensic field.

Acknowledgements We gratefully thank Harald Niederstätter (Institute of Legal Medicine, University of Innsbruck, Austria) for providing the control region sequences and for helpful comments. We also acknowledge the technical advice and valuable discussions of Mike D. Coble, Rebecca S. Hamm (both Armed Forces DNA Identification Laboratory, Rockville, MD) and Peter M. Vallone (National Institute of Standards and Technology, Gaithersburg, MD). This research was supported by a research grant (Nr. 14031/89-02) from the University of Innsbruck, Austria.

References

1. Fakhrai-Rad H, Pourmand N, Ronaghi M (2002) Pyrosequencing: an accurate detection platform for single nucleotide polymorphisms. Hum Mutat 19:479–485
2. Stoneking M (2001) Single nucleotide polymorphisms. From the evolutionary past. Nature 409:821–822
3. Tseng LH, Chen PJ, Lin MT et al. (2002) Simultaneous genotyping of single nucleotide polymorphisms in the IL-1 gene complex by multiplex polymerase chain reaction-restriction fragment length polymorphism. J Immunol Methods 267:151–156
4. Berger B, Holzl G, Oberacher H, Niederstätter H, Huber CG, Parson W (2002) Single nucleotide polymorphism genotyping by on-line liquid chromatography-mass spectrometry in forensic science of the Y-chromosomal locus M9. J Chromatogr B Analyt Technol Biomed Life Sci 782:89–97
5. Gusmao L, Alves C, Costa S et al. (2002) Point mutations in the flanking regions of the Y-chromosome specific STRs DYS391, DYS437 and DYS438. Int J Legal Med 116:322–326
6. Gill P (2001) An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. Int J Legal Med 114:204–210
7. 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
8. Parsons TJ, Coble MD (2001) Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial DNA genome. Croat Med J 42:304–309
9. Morley JM, Bark JE, Evans CE, Perry JG, Hewitt CA, Tully G (1999) Validation of mitochondrial DNA minisequencing for forensic casework. Int J Legal Med 112:241–248
10. Norton N, Williams NM, Williams HJ et al. (2002) Universal, robust, highly quantitative SNP allele frequency measurement in DNA pools. Hum Genet 110:471–478
11. Ambach E, Parson W, Niederstatter H, Budowle B (1997) Austrian Caucasian population data for the quadruplex plus amelogenin: refined mutation rate for HumvWFA31/A. J Forensic Sci 42:1136–1139
12. Santa Lucia J Jr (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc Natl Acad Sci USA 95:1460–1465
13. Schutz E, Ahsen N von (1999) Spreadsheet software for thermodynamic melting point prediction of oligonucleotide hybridization with and without mismatches. Biotechniques 27:1218–1224
14. 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
15. 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
16. Torroni A, Richards M, Macaulay V et al. (2000) mtDNA haplogroups and frequency patterns in Europe. Am J Hum Genet 66:1173–1177
17. Herrnstadt C, Elson JL, Fahy E et al. (2002) Reduced-median-network analysis of complete mitochondrial DNA coding region sequences for the major African, Asian, and European haplogroups. Am J Hum Genet 70:1152–1171
18. Ingman M, Kaessmann H, Paabo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. Nature 408:708–713
19. Legros F, Chatzoglou E, Frachon P et al. (2001) Functional characterization of novel mutations in the human cytochrome b gene. Eur J Hum Genet 9:510–518
20. Torroni A, Huoponen K, Francalacci P et al. (1996) Classification of European mtDNAs from an analysis of three European populations. Genetics 144:1835–1850
21. Wallace DC, Brown MD, Lott MT (1999) Mitochondrial DNA variation in human evolution and disease. Gene 238:211–230
22. Richards M, Corte-Real H, Forster P et al. (1996) Paleolithic and neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet 59:185–203
23. Richards MB, Macaulay VA, Bandelt HJ, Sykes BC (1998) Phylogeography of mitochondrial DNA in western Europe. Ann Hum Genet 62:241–260
24. Lee SD, Lee YS, Lee JB (2002) Polymorphism in the mitochondrial cytochrome B gene in Koreans. An additional marker for individual identification. Int J Legal Med 116:74–78
25. Finnila S, Lehtonen MS, Majamaa K (2001) Phylogenetic network for European mtDNA. Am J Hum Genet 68:1475–1484
26. Maca-Meyer N, Gonzalez AM, Larruga JM, Flores C, Cabrera VM (2001) Major genomic mitochondrial lineages delineate early human expansions. BMC Genet 2:13
27. Richards M, Macaulay V, Hickey E et al. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67:1251–1276
28. Malyarchuk BA, Grzybowski T, Derenko MV, Czarny J, Wozniak M, Miscicka-Sliwka D (2002) Mitochondrial DNA variability in Poles and Russians. Ann Hum Genet 66:261–283
29. Hofmann S, Jaksch M, Bezold R, Mertens S, Aholt S, Paprotta A, Gerbitz KD (1997) Population genetics and disease susceptibility: characterization of central European haplogroups by mtDNA gene mutations, correlation with D loop variants and association with disease. Hum Mol Genet 6:1835–1846
30. Van Der Walt JM, Nicodemus KK, Martin ER et al. (2003) Mitochondrial polymorphisms significantly reduce the risk of Parkinson disease. Am J Hum Genet 72:804–811
31. Gabriel MN, Calloway CD, Reynolds RL, Andelinovic S, Primorac D (2001) Population variation of human mitochondrial DNA hypervariable regions I and II in 105 Croatian individuals demonstrated by immobilized sequence-specific oligonucleotide probe analysis. Croat Med J 42:328–335
32. Lutz S, Weisser HJ, Heizmann J, Pollak S (1997) A third hypervariable region in the human mitochondrial D-loop. Hum Genet 101:384