

Anita Brandstätter¹
 Antonio Salas²
 Harald Niederstätter¹
 Christoph Gassner³
 Angel Carracedo²
 Walther Parson¹

Research Article

Dissection of mitochondrial superhaplogroup H using coding region SNPs

¹Institute of Legal Medicine,
 Innsbruck Medical University,
 Innsbruck, Austria

²Unidad de Genética,
 Instituto de Medicina Legal,
 Facultad de Medicina,
 Santiago de Compostela,
 and Centro Nacional de Genotipado
 (CeGen),
 Hospital Clínico Universitario,
 Galicia, Spain

³Central Institute for Blood
 Transfusion and Immunological
 Department,
 General Hospital and University
 Clinics,
 Innsbruck, Austria

Received October 13, 2005
 Revised November 29, 2005
 Accepted November 29, 2005

Haplogroup H (hg H) includes about 40–50% of the West Eurasian mitochondrial DNA (mtDNA) samples investigated so far. In order to enhance discrimination within this haplogroup we selected 45 coding region SNPs that allow to ascribe samples to the main phylogenetic branches of super hg HV (that embraces hg H) and, in particular, to H sublineages with a much finer resolution than previous studies. SNP selection was carried out using the most up-to-date available literature on population and forensic genetics and extended by means of phylogenetic analysis of complete or coding region genomes (>430) and control region sequences. A meticulous inspection of the H phylogeny led us to the observation of various but uncharacterized subclades of hg H. The selected SNPs were amplified in two PCR-multiplex reactions and subsequently targeted in three single-base extension multiplex reactions. A total of 2214 West Eurasian samples were screened for hg H specific loci 2706 and 7028, of which 859 fell in hg H and were further subjected to subhaplogroup typing. We observed 35 different subhaplogroups in total, 33 of which were found at frequencies below 5%. This assay can be used as a prescreening tool in forensic casework for rapid discrimination between divergent lineages (very effective for high-volume crime cases) or as discriminatory assay, when identical hg H haplotypes were obtained by control region sequencing.

Keywords: Haplogroup H / Minisequencing / Multiplex system / Phylogeny / Subhaplogroups
 DOI 10.1002/elps.200500772

1 Introduction

The application of mitochondrial DNA (mtDNA) analysis in forensic investigations has recently been influenced by the advent of new SNP typing technologies. The latest approaches center on rapid and cost-effective-screening procedures that provide information on the mitochondrial haplogroup affiliation of casework samples [1, 2]. These new technologies play an important role in the preselection of samples deriving from suspects and crime scenes in high-volume cases. The quick elimination of exclusionary samples from further investigations leads to a drastic reduction of the overall

analysis time. In some cases, SNP screening may even provide sufficient information for a preliminary exclusion scenario of samples involved in a crime case and thus supply the contracting authority with a fast and reliable feedback.

mtDNA profiles from West Eurasian populations can be categorized into ten major haplogroups (namely, H, J, K, N1, T, U4, U5, V, X, and W; e.g. [1, 3–10]) by the presence or absence of distinctive polymorphisms residing in the coding region as well as in the control region of the mitochondrial genome. The different haplogroups are far from being evenly distributed in the population; specifically one haplogroup dominates the scenario: haplogroup H (hg H). Approximately 40–50% of samples collected from a West Eurasian population sample belong to this specific haplogroup [1, 8, 9, 11–13]. If, for example, the screening strategy presented in [1] is applied, then this comparatively high percentage of samples can only be further differentiated into hap-

Correspondence: Professor Walther Parson, Institute of Legal Medicine, Innsbruck Medical University, Muellerstrasse 44, A-6020, Innsbruck, Austria

E-mail: walther.parson@i-med.ac.at
Fax: +43-5125072764

Abbreviations: hg H, haplogroup H; mtDNA, mitochondrial DNA

logroups H1 and H3; the SNPs targeted in [2] additionally discriminate five further H-haplogroups (see below for changes in nomenclature). In order to increase the discrimination rate inside hg H without leaving the level of rapid-screening applications, we developed a fast SNP-screening method for a detailed dissection of hg H into 44 different subhaplogroups (including H* and the main branches of the entire super-hg HV). The combination of a haplogroup-determination procedure as described in [1] or [2] and the hg H-screening technology presented here will still lead to a quicker result than the direct sequencing approach.

2 Materials and methods

2.1 DNA samples and extraction

Blood samples were taken from 2214 unrelated and anonymized platelet donors, which corresponded to 0.324% of the total population of Tyrol, Austria. Platelet donors were in the age group 18–65 years and approximately 37% of them were female. Samples may be expected to represent the population of the provincial capital Innsbruck and its proximity best. However, since Innsbruck is of locally exclusive relevance for – to name some – industry, academia, and health care and hence attracts people from all over Tyrol, its population may also show a fairly good representation of the federal state of Tyrol. Total DNA was isolated from EDTA-anticoagulated blood using Nucleon BACC2 reagents (Amersham, Buckinghamshire, UK) and prescreened for hg H by typing the makers 2706 and 7028 in a real-time PCR assay. In total, 859 samples (38.8%) turned out to belong to hg H. Aliquots of those DNA extracts were compiled by automatically dispensing them in a 96-well plate with a Tecan Genesis 150/8 robotic sample processor (Tecan Trading AG, Hombrechtikon, Switzerland). Four wells *per* plate were left empty for no-template controls. This led to a set of nine-and-a-half 96-well plates with DNA extracts belonging to hg H which was then amended to high-throughput genotyping.

2.2 SNP selection

The haplogroup diagnostic sites were selected using data from the literature [2, 5, 7–9, 14, 15]. Basically, the most up-to-date and comprehensive phylogeny provided by the recent contributions of [8] and [9] were partially reconstructed using the data reported by [16]. Following the hierarchical principle for nomenclature due to [11], we here renamed some H subbranches and propose new ones. Here we explain the phylogenetic framework that

underlies the SNP selection. Nucleotide variants are relative to the revised Cambridge reference sequence (rCRS [17, 18]).

The H phylogeny based on the reconstruction of more than 430 coding regions and complete genomes displays a large number of independent basal branches. Apart from the 15 different subhaplogroups named by [8] and [9] (some of them previously reported by [2, 5, 7]), with a number of subclades irradiating from most of the main branches, we describe some new H lineages and sublineages since we found evidences that they are not uncommon in western European populations. H1 is defined by G3010A, which, however, is a (relatively) highly recurrent mutation in the mtDNA coding region. H1a, characterized by control region transitions A73G and A16162G [9], is further subdivided here into H1a1 (defined by T6365C and T16209C), and H1a2, defined by transversion A8271T. Some evidences seem to indicate that site A73G might precede A16162G phylogenetically within H1a, which could necessitate shifting of subclade names. On the other hand [9] anticipated that the motif A73G A16162G may have arisen more than once in hg H. Transition T477C obviously defines a new subbranch of H1, named H1c. It can be further subdivided into H1c1 (defined by A9150G and T16263C) and H1c2 (defined by G8764A and C12858T). We corroborate that H1b, characterized by nucleotide changes T16189C and T16356C [9], is additionally supported by the coding region site A3796G (as indicated in Fig. 1 of [8]). We observed that T8602C (together with T14212C) could define a new subclade, which would be supported by two coding region sequences (#91 and #322) from [7] (note that, unfortunately, no control region data was available in the Herrstadt *et al.* study, which leaves the door open to the existence of additional control region diagnostic sites for this subclade); therefore, and for the time being, we will refer to this seeming cluster as H1–8602. It is important to mention that T8602C may occur twice within H1, as individual #322 in [7] shares a mutation at nucleotide position 6365 with other samples (six complete genomes from [16]) belonging to hg H1a1. We also observe the presence of another potential H1 subbranch, that might be defined by the transition T8473C (shared by four coding region sequences of [7]), here referred to as H1–8473, waiting for further confirmatory data. Finally, we note that the highly variable transition T16093C is present in most of the H1f mtDNAs, a sublineage previously defined by variants T4452C T7309C A9066G T16189C [9].

The rCRS belongs to H2b ([8, 9]; see also [5]). We follow [9] for nomenclature of H2a, defined by G951A, and its main subclade H2a1, defined by the additional transition

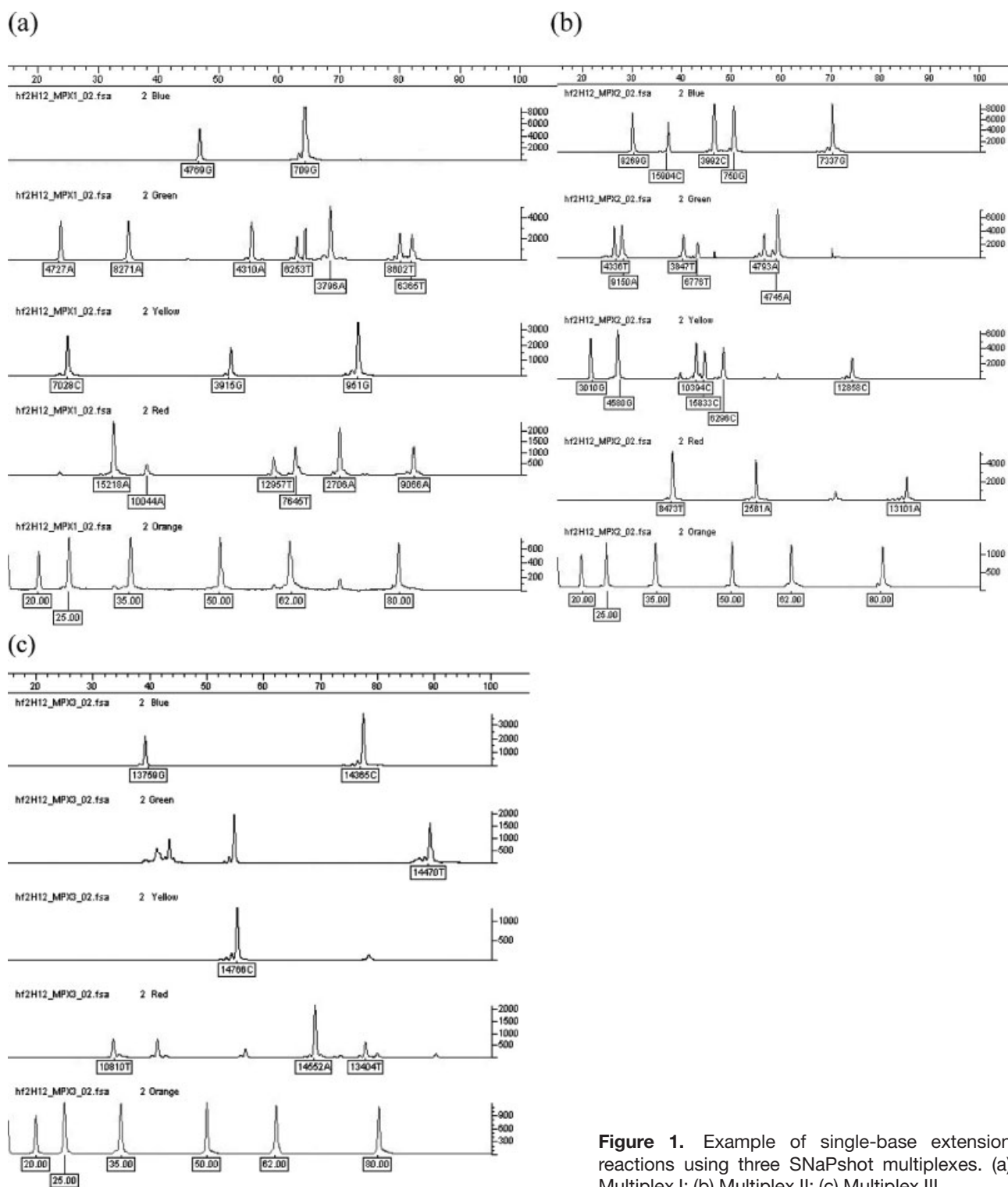


Figure 1. Example of single-base extension reactions using three SNaPshot multiplexes. (a) Multiplex I; (b) Multiplex II; (c) Multiplex III.

C16354T. In agreement with [8], H2c is characterized by T10810C and A16274G. Then, the polymorphism A4769G is carried by all non-H2 sequences.

H3 is determined by site 6776 [7]. Within H3, four coding region sequences from [7] carry the transition T13404C, which seems to point to a new H3 sub-branch, H3a (see

also [8]). This mutation leads to a loss of the *TaqI* restriction site at 13404. We anticipate that transversion C16239G seems to be connected with some subclade of H3a (however, based on a weak evidence: the complete genome #47 from [8]); additional evidence comes from the fact that the HVS-I motif C16239G T16311C has been frequently reported in the literature (*e.g.*, three instances in [19]). We also name here two other sublineages, H3b, determined by nucleotide change A2581G, and H3c, defined by T12957C.

H4 was previously defined by an array of six mutations, namely, C3992T A4024G T5004C G9123A C14365T A14582G (*i.e.* [9]; see also [7] for an earlier definition of H4). Subclade H4a is characterized by position G8269A according to [9]; this branch can be further subdivided by transition A10044G (as anticipated by [9]) defining H4a1. The later includes a subclade, H4a1a, characterized by the transition A73G (and the hypervariable deletion 523–524del).

H5 is defined by C456T and T16304C [9]. Within H5, a mutational change T4336C characterizes H5a, and the additional transition C15833T defines H5a1 [9].

H6 (characterized by sites 239, 16362, and 16482 [8, 9]) has two main sub-branches, H6a (G3915A; defined by [2] as H6 and renamed by [9]), and H6b (A16300G; [9]). Within H6a, site 9380 determines H6a1 [9]; and we confirm the additional site 4727 to its motif. As evidenced from the sequence #H3–09 of [16] and #022 of [20] transition at np 3915 occurs also outside H6 within hg H (and is also recurrent in non-H haplogroups). Therefore, these two samples and #312 and #348 from [7], which all harbor the motif 3915–6296 likely form another basal subclade within H, here tentatively named H17. It is possible that the sequences H3–05 and H3–06 of [16] (and #150, #333 from [7]) are also related to this subclade, as all these full sequences share the G3915A – G16129A motif and they do not carry other mutations characteristic to H6.

Transition A4793G is diagnostic of hg H7, as in [9].

H8 is characterized by sites A13101C, T16288C, and T16362C according to [8, 9]; we observe that 709 could be added to its diagnostic motif (note however that this site is phylogenetically relatively unstable).

We follow the nomenclature of [8] for subhaplogroups: H9 (G3591A A4310G T13020C; replacing the early H9 definition of [9]), H10 (T14470A), H12 (C3936T A14552G), H14 (T7645C G11377A), and H15 (T55C T57C T6253C), and other minor branches of these subclades. For hg H13, we supplement with the information taken from [10]; then we rename H13 [8] as the clade defined by C2259T C14872T, whereas H13a would cluster those sequences carrying

the additional sites 4745 13680. We then rename H13a [8] as H13a1 to the subclade defined by the extra site 7337. The definition of H11 (T8448C G13759A T16311C) and H11a (T961G and A16293G on top) is as suggested by [8]; a branching pattern supported on the data provided by [20–22].

We here define H16, characterized by C10394T, which is supported by seven complete genomes from [16]. Those mtDNAs carrying G709A, but without any of the diagnostic sites that define H1 to H16, will be referred to as H-709 (see also comments above concerning H8). In addition, other SNPs were chosen because they allow to discriminate the main branches of the entire super-hg HV: T3847C (within an array of other mutations [8, 10] defines (pre-HV)1 [13]), C15904T that precedes haplogroup pre*V2 and V [8, 23], G4580A (haplogroup V) [8, 23], T14766C (hg HV) [12], and A15218G, that together with two other mutations defines hg HV1 (HV1 was first defined by [13] and the placement of A15218G in the tree is by [8]). Finally, note that for most (sub)haplogroups, several SNPs are diagnostic, which allowed for alternative locus selection for complex primer design scenarios. An overview of the selected sites and their associated haplogroups is given in Table 1.

2.3 Primers: design and evaluation

Primers for amplification and minisequencing were designed using the program VisualOMP (DNA Software, Ann Arbor, MI, USA). Desalted and HPLC-purified oligonucleotides were purchased from Proligo (Boulder, CO, USA); primer integrity was monitored using LC ESI MS [24]. Minisequencing primers were tested experimentally in a “selfextension reaction” using the Primer Focus Kit (AB – Applied Biosystems, Foster City, CA, USA).

2.4 PCR-multiplex amplification

The target sequences were amplified in a 13-plex and a 3-plex PCR reaction (Table 1). For amplification, to each well of a MicroAmp Optical 96-well reaction plate (AB), 23 μ L of PCR-master mix, 2.0 units of Advantage 2 (Clontech, BD Biosciences, Palo Alto and Mountain View, CA, USA), 1.0 unit of reaction buffer (Clontech), 300 μ M each dNTP (AB), 6.25 μ g BSA, 5 mM MgCl₂, (amplification primers and their concentrations are specified in Table 2) were added. DNA extract (2 μ L; containing ~4 ng of genomic DNA) was then put into the PCR-master mix with a multichannel pipette. Thermal cycling was carried out on a GeneAmp 9700 (AB) or on an iCycler (BioRad, Hercules, CA, USA) using the following conditions: 95°C for

Table 1. Selected SNP sites, their associated haplogroups and the concept of multiplexing on the levels of PCR amplification and SNaPshot minisequencing

SNP marker	Haplogroup	Target	Target positions	PCR-multiplex	SNP (ancestral-derived)	SNaPshot-Multiplex
709	H-709	hPCR01	576– 1 030	13-plex	G–A	MPX1
750	H2b	hPCR01	576– 1 030	13-plex	A–G	MPX2
951	H2a	hPCR01	576– 1 030	13-plex	G–A	MPX1
2 581	H3b	hPCR02	2 435– 3 127	13-plex	A–G	MPX2
2 706	H	hPCR02	2 435– 3 127	13-plex	A–G	MPX1
3 010	H1	hPCR02	2 435– 3 127	13-plex	G–A	MPX2
3 796	H1b	hPCR03	3 608– 4 167	13-plex	A–G	MPX1
3 847	(pre-HV)1	hPCR03	3 608– 4 167	13-plex	T–C	MPX2
3 915	H6a	hPCR03	3 608– 4 167	13-plex	G–A	MPX1
3 992	H4	hPCR03	3 608– 4 167	13-plex	C–T	MPX2
4 310	H9	hPCR04	4 196– 4 881	13-plex	A–G	MPX1
4 336	H5a	hPCR04	4 196– 4 881	13-plex	T–C	MPX2
4 580	V	hPCR04	4 196– 4 881	13-plex	G–A	MPX2
4 727	H6a1	hPCR04	4 196– 4 881	13-plex	A–G	MPX1
4 745	H13a	hPCR04	4 196– 4 881	13-plex	A–G	MPX2
4 769	H2	hPCR04	4 196– 4 881	13-plex	A–G	MPX1
4 793	H7	hPCR04	4 196– 4 881	13-plex	A–G	MPX2
6 253	H15	hPCR05	6 118– 6 541	13-plex	T–C	MPX1
6 296	H17	hPCR05	6 118– 6 541	13-plex	C–T	MPX2
6 365	H1a1	hPCR05	6 118– 6 541	13-plex	T–C	MPX1
6 776	H3	hPCR06	6 648– 7 159	13-plex	T–C	MPX2
7 028	H	hPCR06	6 648– 7 159	13-plex	C–T	MPX1
7 337	H13a1	hPCR07	7 179– 7 799	13-plex	G–A	MPX2
7 645	H14	hPCR07	7 179– 7 799	13-plex	T–C	MPX1
8 269	H4a1	hPCR08	8 105– 8 782	13-plex	G–A	MPX2
8271	H1a2	hPCR08	8 105– 8 782	13-plex	A–T	MPX1
8 473	H1–8473	hPCR08	8 105– 8 782	13-plex	T–C	MPX2
8 602	H1–8602	hPCR08	8 105– 8 782	13-plex	T–C	MPX1
9 066	H1f	hPCR09	8 946– 9 294	13-plex	A–G	MPX1
9 150	H1c1	hPCR09	8 946– 9 294	13-plex	A–G	MPX2
10 044	H4a1	hPCR10	9 922–10 522	13-plex	A–G	MPX1
10 394	H16	hPCR10	9 922–10 522	13-plex	C–T	MPX2
10 810	H2c	hPCR11	10 649–11 120	3-plex	T–C	MPX3
12 858	H1c2	hPCR12	12 670–13 259	13-plex	C–T	MPX2
12 957	H3c	hPCR12	12 670–13 259	13-plex	T–C	MPX1
13 101	H8	hPCR12	12 670–13 259	13-plex	A–C	MPX2
13 404	H3a	hPCR13	13 320–13 913	3-plex	T–C	MPX3
13 759	H11	hPCR13	13 320–13 913	3-plex	G–A	MPX3
14 365	H4	hPCR14	14 282–14 949	3-plex	C–T	MPX3
14 470	H10	hPCR14	14 282–14 949	3-plex	T–A	MPX3
14 552	H12	hPCR14	14 282–14 949	3-plex	A–G	MPX3
14 766	HV*	hPCR14	14 282–14 949	3-plex	T–C	MPX3
15 218	HV1	hPCR15	15 092–15 486	13-plex	A–G	MPX1
15 833	H5a1	hPCR16	15 726–16 084	13-plex	C–T	MPX2
15 904	pre*V2 & V	hPCR16	15 726–16 084	13-plex	C–T	MPX2

2 min; three cycles of (95°C for 15 s, 52°C for 55 s, and 72°C for 1 min); 20 cycles of (95°C for 15 s, 52°C + 0.4°C per cycle for 55 s, 72°C for 1 min); 12 cycles of (95°C for 15 s, 60°C for 55 s, and 72°C for 1 min); 72°C for 10 min; 4°C until removed from thermocycler. PCR primers and

unincorporated dNTPs were removed by adding 2 µL of ExoSAP-IT (USB Corporation, Cleveland, OH, USA) to each 5 µL PCR reaction. Reactions were incubated at 37°C for 90 min and then at 80°C for 20 min for enzyme deactivation.

Table 2. Primer sequences and concentrations for PCR amplification

PCR	Forward primer	Reverse primer	Final primer concentration in reaction mix
hPCR01	AGTTTATGTAGCTTACCTCCTCAAAGCAAT	CACTTTCGTAGTCTATTTTGTGTCAACTGG	128 nM
hPCR02	GGCATGCTCATAAGGAAAGGTTAAAA AAAGTAA	CGTACAGGGGAGGAATTTGAAGGTAG	256 nM
hPCR03	GCCTCCTATTTATTCTAGCCACCTCTAG	GTGTATGAGTTGGTCGTAGCGGAA	256 nM
hPCR04	CCCTAGCATTACTTATATGATATGTCTCC ATACCCATTACAATCTCC	GGGCTAGTTTTTGTTCATGTGAGAAGA AGCAG	513 nM
hPCR05	CCATCATAATCGGAGGCTTTGGC	CGGTCTGTTAGTAGTATAGTGATGCCAG	256 nM
hPCR06	CCAGGCTTCGGAATAATCTCCCATAT	ATGAATATGATAGTGAAATGGATTTTGG CGTAG	256 nM
hPCR07	TTCCCACAACACTTTCTCGGCC	TGATGGCGGGCAGGATAGTTCAGA	256 nM
hPCR08	GCAATCCCAGGACGTCTAAACCAAA	CGAGGAGGTTAGTTGTGGCAATAAAAAT	128 nM
hPCR09	ACTAGTTATTATCGAAACCATCAGCCTACTC	CGGAGGTCATTAGGAGGGCTGA	256 nM
hPCR10	CCTGATACTGGCATTGTGTAGATGTGG	CCTAGAAGTGAGATGGTAAATGCTAGTATA ATATTTATG	1026 nM
hPCR11	TATTGCCATACTAGTCTTTGCCGCC	AGATATAAAATATGATTAGTTCTGTGGCT GTGAATGTTAT	286 nM
hPCR12	CCAAACATTAATCAGTTCTTCAAATAT CTACTCATCTT	CTTGAAGTGGAGAAGGCTACGATTTTTTT	205 nM
hPCR13	CATCTGTACCCACGCCTTCTTCAA	AGTATGTTGGAGAAATAAAATGTGCATA GTGG	143 nM
hPCR14	CCCTCTCCTTCATAAATTATTCAGCTTCCT	TGGGCGATTGATGAAAAGGCG	571 nM
hPCR15	GGCATTATCCTCCTGCTTGCAACTAT	GGTGAGAATAGTGTTAATGTCATTAAG GAGAGA	205 nM
hPCR16	TCCTAGCCGCAGACCTCCTC	CGGTTGTTGATGGGTGAGTCAATAC	256 nM

2.5 Minisequencing multiplexes using SNaPshot

The minisequencing reactions were carried out in three different multiplex assays (Table 1). The extension primers (Table 3) were premixed for each multiplex master mix. The multiplex primer extension reactions were carried out in a total volume of 10 μ L comprising 2.5 μ L of the SNaPshot™ Multiplex Ready Reaction Mix (AB), 0.5 μ L of 10 \times AmpliTaq Gold® PCR buffer (AB), 1.0 μ L of PCR product, pooled extension primers (MPX1: 3.6 μ L; MPX2: 4.0 μ L; MPX3: 1.8 μ L), and water up to 10 μ L. Thermal cycling and post-extension treatment were conducted following the manufacturer's protocol. Unincorporated ddNTPs were removed with SAP (Shrimp Alkaline Phosphatase; USB).

2.6 Electrophoresis and data analysis

For preparation of samples for electrophoresis, 2 μ L of SNaPshot product was diluted in 50 μ L of Hi-Di form-

amide, and subsequently mixed with 8 μ L of GeneScan-120 LIZ size standard – Hi-Di formamide mixture containing 0.8 μ L of GeneScan-120 LIZ size standard *per* 100 μ L of Hi-Di formamide. Samples were denatured at 95°C for 5 min and then immediately put on ice. Electrophoresis was run on an ABI 3100 Genetic Analyzer using 3100-POP-6™ 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 3.6 NT (all AB) using a self-written macro.

2.7 Random matching probability

The probability of a chance match (M) between two unrelated individuals was calculated as:

$$P(M) = \frac{1}{n^2} \cdot \sum_{i=1}^m x_i^2 \quad [25]$$

Table 3. Extension primers for three multiplex SNaPshot assays

SNP	Extension primer with poly-T tail for spacing	Strand	Length	Concentration
709	(T) ₂₇ TATTAGCTCTTAGTAAGATTACACATGCAAGCATCCCC	F	65	0.10 μM
750	(T) ₁₇ GAGTTCACCCTCTAAATCACCACGATCAAAAAGG	F	50	0.10 μM
951	(T) ₃₂ AACTCAGGTGAGTTTTAGCTTTATTGGGGAGGGGGTGAT	R	71	0.10 μM
2 581	(T) ₁₇ TTTAAGGAACAAGTGATTATGCTACCTTTGCACGGT	R	53	0.05 μM
2 706	(T) ₂₉ TAAATTAAGCTCCATAGGGTCTTCTCGTCTTGCTGTGT	R	68	0.08 μM
3 010	TTAATAGCGGCTGCACCAT	R	20	0.24 μM
3 796	(T) ₂₉ CTATCAACACTACTAATAAGTGGCTCCTTTAACCTCTCC	F	68	0.07 μM
3 847	(T) ₁₃ GTGGAGATAAATCATATTATGGCCA	R	38	0.39 μM
3 915	(T) ₁₄ GTATTCGATGTTGAAGCCTGAGACTAGTTCGGACTC	R	50	0.39 μM
3 992	(T) ₁₄ GTGGTGAGGGTGTATTATAATAATGTTT	R	44	0.19 μM
4 310	(T) ₁₉ ATAAAAGAGTTACTTTGATAGAGTAAATAATAGG	F	53	0.39 μM
4 336	(T) ₂ AGGGATGGGTTGATTCTCAT	R	23	0.04 μM
4 580	TTGGTTAGAACTGGAATAAAAAGCTAG	R	26	0.29 μM
4 727	(T) ₂ CTCCGGACAATGAACCAT	F	20	0.39 μM
4 745	(T) ₂₄ TCCGGACAATGAACCATAACCAATACTACCAATCA	F	59	0.15 μM
4 769	(T) ₁₅ ACCAATCAATACTCATCATTATAATCATAAT	F	47	0.29 μM
4 793	(T) ₂₁ ATAATCATAATAGCTATAGCAATAAACTAGGAAT	F	56	0.39 μM
6 253	(T) ₂₆ TAGACTGTTCAACCTGTTCTGCTCCGGCCTCCACT	R	62	0.04 μM
6 296	(T) ₁₂ GCCGGAGCAGGAACAGGTTGAACAGTCTACCCTCC	F	47	0.19 μM
6 365	(T) ₄₀ TTGTTGTGATGAAATTGATGGCCCCTAAGATAGAGGAGAC	R	80	0.08 μM
6 776	(T) ₁₃ TGTCTACGTCTATTCTACTGTAAATAT	R	41	0.07 μM
7 028	(T) ₂ CGACACGTACTACGTTGTAGC	F	23	0.19 μM
7 337	(T) ₃₁ TAATATTAATAATTTTCATGATTTGAGAAGCCTTCGCTTC	F	71	0.39 μM
7 645	(T) ₂₅ GTCTACAAGACGCTACTTCCCCTATCATAGAAGAGCT	F	62	0.29 μM
8 269	(T) ₉ GGGCCCGTATTTACCCTATA	F	29	0.15 μM
8 271	(T) ₁₀ GGGCCCGTATTTACCCTATAGC	F	32	0.07 μM
8 473	(T) ₈ AAAAATATTAACACAACTACCACCTACC	F	38	0.19 μM
8 602	(T) ₃₇ TTGATGAGATATTTGGAGGTGGGGATCAATAGAGGGGGAA	R	77	0.09 μM
9 066	(T) ₄₀ TGAAGATGATAAGTGTAGAGGGAAGGTTAATGGTTGATAT	R	80	0.39 μM
9 150	(T) ₆ CTAGAAATCGCTGTGCGCTT	F	26	0.29 μM
10 044	(T) ₅ TTAAGGCGAAGTTTATTACTCTTTTTTGAA	R	35	0.39 μM
10 394	(T) ₁₀ TGGCCTATGAGTGACTACAAAAGGATTAGA	F	41	0.24 μM
10 810	(T) ₄ CAATTATATTACTACCACTGACATGACT	F	32	0.52 μM
12 858	(T) ₃₄ GAGCAGATGCCAACACAGCAGCCATTCAAGCAATCCTATA	F	74	0.14 μM
12 957	(T) ₂₂ CATGAGACCCACAACAAATAGCCCTTCTAAACGCTAA	F	59	0.12 μM
13 101	(T) ₅₆ GTGGAAGCGGATGAGTAAGAAGATTCC	R	83	0.16 μM
13 404	(T) ₃₇ CCGGGTCCATCATCCACAACCTTAACAATGAACAAGATAT	F	77	0.16 μM
13 759	(T) ₅ TGCAGGATTTCTCATTACTAACAACATTTCCCCC	F	40	0.13 μM
14 365	(T) ₃₄ TGAGTGTTTTAGTGGGGTTAGCGATGGAGGTAGGATTGGT	R	74	0.20 μM
14 470	(T) ₅₃ AATTTATTTAGGGGGAATGATGGTTGTCTTTGG	R	86	0.26 μM
14 552	(T) ₂₆ GGGGTTTAGTATTGATTGTTAGCGGTGTGGTCCGGTGTG	R	65	0.26 μM
14 766	(T) ₁₉ TCAACTACAAGAACACCAATGACCCCAATACGCAAAA	F	56	0.26 μM
15 218	(T) ₂ CTCCTCAGATTCATTGAAGTCTG	R	29	0.15 μM
15 833	(T) ₉ GACAAGTAGCATCCGTAATACTTACACAACAATC	F	44	0.24 μM
15 904	(T) ₁₁ CCGTTTACAAGACTGGTGTATTA	R	35	0.08 μM

Final primer concentration in the reaction mix and the orientation of each extension primer is given (F: primer binds to L-strand; R: primer binds to H-strand).

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 as observed in the database.

3 Results and discussion

3.1 Multiplexing efforts

The presented hg H resolving strategy was conceived as a two-stage multiplex method with the intention to amplify all PCR fragments in a single 16-plex reaction and to target the 45 SNP sites in two minisequencing multiplex reactions. Since SNP interrogation using primer extension does not permit any flexibility with respect to the location of the 3'-end of the primer, and the choice of SNP sites was restricted phylogenetically, some primers had to be selected although their design was not perfect in terms of extendable dimer formation between primers. Seven sites did not fit in any of the two conceived SNaPshot multiplexes: 10810, 13404, 13759, 14365, 14470, 14552, and 14766 defining haplogroups H2c, H3a, H11, H4a, H10, H12, and HV*, respectively. So, all the 859 samples were first amplified using a 13-plex and typed with the minisequencing multiplexes MPX1 and MPX2 (Table 1). After assigning those samples to their respective haplogroups, samples that could still not be differentiated from hg H* and samples that were assigned to haplogroups H3* (excluding H3b and H3c) and H4 were amplified with the 3-plex (amplifying PCR fragments hPCR11, hPCR13, and hPCR14) and typed with the minisequencing MPX3 (targeting the seven sites described above).

3.2 SNP genotyping

All the 859 samples were successfully amplified with the 13-plex and typed with the minisequencing multiplexes MPX1 and MPX2, giving clear results for all the 38 typed markers (Figs. 1a and b). Among the 859 samples, 366 were amplified with the 3-plex and typed with the minisequencing multiplex MPX3 (Fig. 1c). The results of the SNaPshot reactions were confirmed by direct sequencing of the PCR products using the control DNA 9947A (Promega Corporation, Madison, WI, USA). Macroassisted evaluation of the raw data (with a macro written for Genotyper software (AB)) enabled a very fast and unambiguous hg-assessment of the individual profiles.

Within MPX2, two small extraneous peaks were observed in nearly all cases: in the green channel, in many cases there was a peak sized ~45 bases, and in the yellow channel an artefact peak sized ~39.5 bases occurred frequently. Within MPX3, four extraneous peaks were found in all electropherograms: in the green channel, artefact peaks occurred in the region of bases sized

Table 4. Frequencies of the different H (sub)haplogroups

Haplogroup	n	Frequency (%)
H*	283	32.9
H1*	163	19.0
H1a1	3	0.3
H1b	24	2.8
H1c1	16	1.9
H1c2	6	0.7
H1f	2	0.2
H1-8602	2	0.2
H1-8473	2	0.2
H2*	12	1.5
H2a	22	2.6
H2b	21	2.4
H3*	19	2.2
H3a	39	4.6
H3b	8	0.9
H3c	1	0.1
H4*	14	1.6
H4a	2	0.2
H4a1	4	0.5
H5a	25	2.9
H5a1	17	2.0
H6a	4	0.5
H6a1	21	2.4
H7*	20	2.3
H8*	1	0.1
H10	31	3.6
H11	11	1.4
H13a	9	1.0
H13a1	18	2.1
H14	7	0.8
H15	17	2.0
H16	17	2.0
H17	1	0.1
H-709	17	2.0

40–45 and another high artefact peak appeared at base size position 55.8. In the red channel, artefact peaks sized ~41.2 and ~55.6 bases appeared in nearly all electropherograms. We identified the sources for these extraneous peaks by single-plex analysis of each of the SNPs and were able to assign them, respectively. None of these artefacts affected the SNP interrogation process, as they did not coincide with category ranges of expected loci.

3.3 Frequencies of the different H (sub)haplogroups

In order to specifically type samples belonging to hg H, a population sample of 2214 individuals was screened for the presence or absence of the A2706G and C7028T nucleotide substitutions with a real-time PCR assay. West Eurasian mtDNA lineages lacking the transitions at both

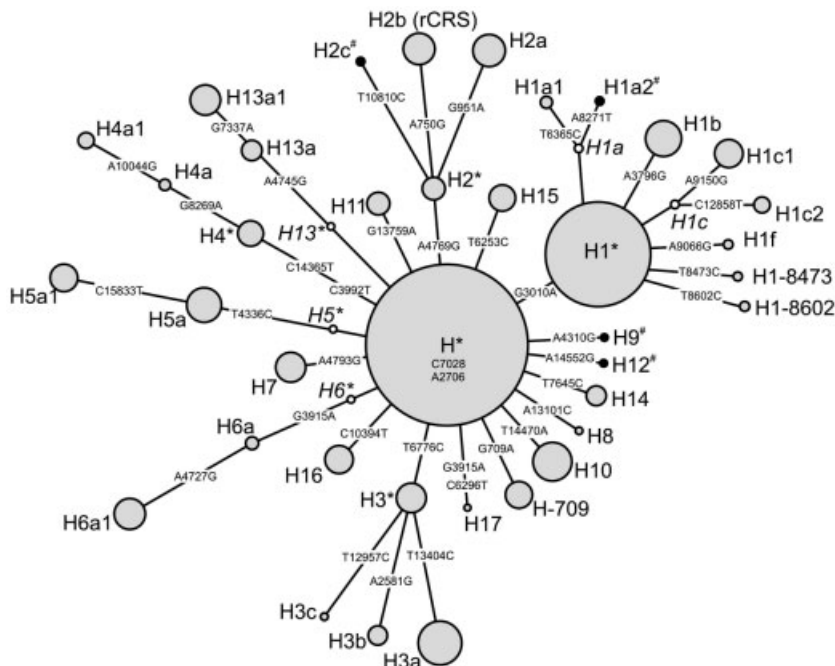


Figure 2. Median joining (MJ) network of hg HV. Node areas correspond to subhaplogroup frequencies for the 859 samples analyzed. Running of MJ algorithm and drawing of the resulting graph was performed with the computer program Network 4.1.1.1 (A. Röhl; © Fluxus Technology; Fluxus Engineering Web site: <http://www.fluxus-engineering.com>). Haplogroups that were not observed (but their diagnostic SNP were interrogated) in this West-Eurasian population sample are denoted with hash marks, the respective nodes are entirely black. “Nontested” hg H sub-branches (e.g., H1c, H6*, etc.) are also displayed in the figure (indicated as empty nodes and italicized haplogroup label) according to the phylogenetic description in Section 2. Figure shows only the SNPs genotyped in the present SNaPshot design.

sites generally fall into hg H [7–9]. In total, 859 samples lacked both the A2706G and C7028T polymorphisms and were thus subjected to a further hg H subdivision.

A total of 35 different subhaplogroups were found in 859 samples belonging to hg H. The majority of samples (67.1%) was resolved into smaller subhaplogroups of hg H (Table 4). hg H 1 encompassed 25.3% of the hg H samples, and a further subdivision could be established by the presence of polymorphisms T6365C (H1a1; 0.3%), A3796G (H1b; 2.8%), A9150G (H1c1; 1.9%), C12858T (H1c2; 0.7%), A9066G (H1f; 0.2%), T8602C (H1–8602; 0.2%), and T8473C (H1–8473, 0.2%). The remaining 33 haplogroups occurred at frequencies below 5% (Fig. 2). Four haplogroups that could be distinguished with the selected SNP sites were not found in this West-Eurasian population sample: H1a2, H2c, H9, and H12.

3.4 Random matching probability

The most common haplotype found was a profile exhibiting two differences to the rCRS (750G and 4769G), thus assigned to hg H* (32.9%), as the rCRS lies within cluster H2b. Overall, 35 different lineages were found in 859 individuals, and three of them appeared only once in the dataset (H3c, H8*, and H17). The probability of a random match between two unrelated hg H individuals was calculated as 15.6%, corresponding to a power of discrimination of 84.4%, which adds substantial discrimination power to the SNP analysis developed in [1].

4 Concluding remarks

Analysis of mtDNA has found a vital niche in forensic applications, where autosomal STR markers fail to give conclusive results due to the low amount of intact nuclear DNA. However, even if the external circumstances of the stains recovered from a crime scene favor mtDNA analysis, it is frequently not being applied because of the high efforts that are involved with control region sequencing.

For this reason rapid-screening procedures are of high interest, as they enable a preselection and thus reduction of samples that need to be sequenced later on. In some cases they can even quickly lead to exclusion scenarios and thus allow rapid communication with the authorities involved in the case. In addition, SNP interrogation can provide additional discrimination of otherwise identical control region sequences [16, 26–28]. The SNP markers selected for this study provide a very fine resolution within hg H. Two major groups (H* and H1) however still remain to be resolved by the application of appropriate sites. The presented hg H dissecting multiplex system serves the forensic investigator in two important ways: as a screening tool, it provides quick information for the elimination of exclusionary samples; in addition, the strategy supplies further discrimination of samples that cannot be discriminated with control region sequencing. Thus, the presented hg H dissecting multiplex system poses a new and innovative tool for forensic case-work, which becomes very effective in investigations involving West Eurasian populations.

We would like to thank Burkhard Berger (Institute of Legal Medicine, Innsbruck Medical University, Austria) for helpful discussions and sample compilation. We are grateful to Herbert Oberacher and Florian Pitterl (Institute of Legal Medicine, Innsbruck Medical University, Austria) for monitoring the quality of the synthesized primers by means of mass spectrometry. We would like to thank Hans Jürgen-Bandelt for helpful discussion on nomenclature and phylogeny and two anonymous reviewers for their useful comments on the manuscript. This work was supported by a grant from the European Community (GROWTH programme, SNPforID project, contract G6RD-CT-2002-00844).

5 References

- [1] Brandstätter, A., Parsons, T. J., Parson, W., *Int. J. Legal Med.* 2003, 117, 291–298.
- [2] Quintáns, B., Álvarez-Iglesias, V., Salas, A., Phillips, C. et al., *Forensic Sci. Int.* 2004, 140, 251–257.
- [3] Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M. et al., *Genetics* 1996, 144, 1835–1850.
- [4] Wallace, D. C., Brown, M. D., Lott, M. T., *Gene* 1999, 238, 211–230.
- [5] Finnilä, S., Lehtonen, M. S., Majamaa, K., *Am. J. Hum. Genet.* 2001, 68, 1475–1484.
- [6] Allard, M. W., Miller, K., Wilson, M., Monson, K. L. et al., *J. Forensic Sci.* 2002, 47, 1215–1223.
- [7] Herrnstadt, C., Elson, J. L., Fahy, E., Preston, G. et al., *Am. J. Hum. Genet.* 2002, 70, 1152–1171.
- [8] Achilli, A., Rengo, C., Magri, C., Battaglia, V. et al., *Am. J. Hum. Genet.* 2004, 75, 910–918.
- [9] Loogväli, E.-L., Roostalu, U., Malyarchuk, B. A., Derenko, M. V. et al., *Mol. Biol. Evol.* 2004, 21, 2012–2021.
- [10] Palanichamy, M. G., Sun, C., Agrawal, S., Bandelt, H.-J. et al., *Am. J. Hum. Genet.* 2004, 75, 966–978.
- [11] Richards, M., Macaulay, V., Bandelt, H.-J., Sykes, B. C., *Ann. Hum. Genet.* 1998, 62, 241–260.
- [12] Macaulay, V., Richards, M., Hickey, E., Vega, E. et al., *Am. J. Hum. Genet.* 1999, 64, 232–249.
- [13] Richards, M., Macaulay, V., Hickey, E., Vega, E. et al., *Am. J. Hum. Genet.* 2000, 67, 1251–1276.
- [14] Ingman, M., Kaessmann, H., Pääbo, S., Gyllensten, U., *Nature* 2000, 408, 708–713.
- [15] Herrnstadt, C., Preston, G., Howell, N., *Am. J. Hum. Genet.* 2003, 72, 1585–1586.
- [16] Coble, M. D., Just, R. S., O’Callaghan, J. E., Letmanyi, I. H. et al., *Int. J. Legal Med.* 2004, 118, 137–146.
- [17] Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. et al., *Nature* 1981, 290, 457–465.
- [18] Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N. et al., *Nat. Genet.* 1999, 23, 147–147.
- [19] Helgason, A., Hickey, E., Goodacre, S., Bosnes, V. et al., *Am. J. Hum. Genet.* 2001, 68, 723–737.
- [20] Howell, N., Smejkal, C. B., Mackey, D. A., Chinnery, P. F. et al., *Am. J. Hum. Genet.* 2003, 72, 659–670.
- [21] Rieder, M. J., Taylor, S. L., Tobe, V. O., Nickerson, D. A., *Nucleic Acids Res.* 1998, 26, 967–973.
- [22] Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V. et al., *Proc. Natl. Acad. Sci. USA* 2003, 100, 171–176.
- [23] Torroni, A., Bandelt, H.-J., Macaulay, V., Richards, M. et al., *Am. J. Hum. Genet.* 2001, 69, 844–852.
- [24] Oberacher, H., Niederstätter, H., Parson, W., *J. Mass Spectrom.* 2005, 40, 932–945.
- [25] Stoneking, M., Hedgecock, D., Higuchi, R. G., Vigilant, L. et al., *Am. J. Hum. Genet.* 1991, 48, 370–382.
- [26] Just, R. S., Irwin, J. A., O’Callaghan, J. E., Saunier, J. L. et al., *Forensic Sci. Int.* 2004, 146, 147–149.
- [27] Vallone, P. M., Just, R. S., Coble, M. D., Butler, J. M. et al., *Int. J. Legal Med.* 2004, 118, 147–157.
- [28] Pereira, L., Richards, M., Goios, A., Alonso, A. et al., *Forensic Sci. Int.* 2006, 159, 43–50.