Int J Legal Med (2009) 123:357–360
DOI 10.1007/s00414-009-0350-y

SHORT COMMUNICATION

# Subtyping of Y-chromosomal haplogroup E-M78 (E1b1b1a) by SNP assay and its forensic application

**S. Caratti · S. Gino · C. Torre · C. Robino**

Received: 13 November 2008 / Accepted: 22 April 2009 / Published online: 9 May 2009


**Abstract** The continual discovery of new single-nucleotide polymorphisms (SNPs) has led to an increased resolution of the Y chromosome phylogeny. Some of these Y-SNPs have shown to be restricted to small geographical regions and therefore may prove useful in the forensic field as tools for the prediction of population of origin of unknown casework samples. Here, we describe a system for the molecular dissection of haplogroup E-M78 (E1b1b1a), consisting of multiplex polymerase chain reaction and minisequencing of M78 and nine population-informative Y-SNPs (M148, M224, V12, V13, V19, V22, V27, V32, V65) in a single reaction. Sensitivity and admixture studies demonstrated that the SNP protocol allows robust genotyping from as little as 50 pg of male DNA, even in the presence of 500-fold amounts of female DNA. In order to evaluate the suitability of E1b1b1a, subhaplogrouping for population-of-origin prediction, the distribution of E-M78 and its derived variants was determined in an Italian population sample ($n$=326).

**Keywords** Forensic genetics · Y chromosome · Single-nucleotide polymorphism · Haplogroup · Minisequencing

**Electronic supplementary material** The online version of this article (doi:10.1007/s00414-009-0350-y) contains supplementary material, which is available to authorized users.

S. Caratti · S. Gino · C. Torre · C. Robino (✉)
Department of Anatomy, Pharmacology,
and Legal Medicine, University of Turin,
Corso Galileo Galilei 22,
10126 Turin, Italy
e-mail: carlo.robino@unito.it

## Introduction

Analysis of Y-chromosomal polymorphisms provides an extremely useful tool in forensic DNA testing. Y-chromosomal short tandem repeat (STR) haplotypes can be typed in samples from sexual assault cases, even in the presence of a high female DNA background [1].

Y chromosome haplogroups, defined by the combination of allelic states at binary slowly evolving single-nucleotide polymorphism (SNP) loci, are distributed nonrandomly among human populations [2] and may allow prediction of the geographic or ethnic origin of unknown casework samples [3]. Because of their simple molecular structure, large sets of SNPs are amenable to multiplexing and high-throughput analysis, and genotyping can be done in much smaller amplicons than those of conventional STRs [4]. SNP markers thus provide an important instrument for rapid exclusion in forensic mass screening [5, 6] and genotyping of highly degraded DNA in disaster victim identification [7].

Among the available SNP genotyping assays, multiplex polymerase chain reaction (PCR) amplification followed by single base extension (SBE) using the SnaPshot mini-sequencing technology is at present the most popular method in forensic laboratories [8]. Several protocols for the simultaneous detection of Y-SNPs defining the major clades of the Y chromosome phylogenetic tree have been recently described for evolutionary and forensic purposes [9–11]. Although marked differences in the global distribution of these major haplogroups are seen between distinct geographic areas, very few of them are entirely confined to single populations [12, 13]. Therefore, integration of these Y-SNP sets with newly discovered binary polymorphisms [14], some of which have shown to be restricted to small geographical regions and population specific, is highly

desirable. For instance, Cruciani et al. recently described eight new Y-SNPs (V12, V13, V19, V22, V27, V32, V36, V65) that allow the molecular dissection of haplogroup E-M78 (E1b1b1a), one of the three main subclades of haplogroup E-M35 (E1b1b1; Fig. S1 of Electronic supplementary material) [15, 16]. Haplogroup E-M78 is distributed over a wide area stretching from eastern (21.5%) and northern (18.5%) Africa to the Near East (5.8%) and Europe (7.2%), where it represents the most common E1b1b subhaplogroup [17]. However, many of the newly described subhaplogroups of E-M78 have a much more limited geographical distribution. E-V65 and the paragroups E-M78* and E-V12* are seen mainly in northern Africa; E-V13 is found in Europe, mostly in the Balkan peninsula, while E-V32 is present at high frequencies only in eastern Africa. E- V22 is the only haplogroup showing a wide geographic distribution, including northeastern and eastern Africa, Europe, and western Asia up to southern Asia (Fig. S2 of Electronic supplementary material) [18].

The aim of this study was to develop a minisequencing assay able to subtype in a single reaction all the Y-SNPs that define geographically informative E-M78 subhaplogroups. In order to assess its suitability in forensic casework, the efficiency of the method was tested both in low-copy-number (LCN) DNA conditions and in male/female mixed samples reproducing sexual assault cases. With regards to the inference of geographic origin of unknown stains, the distribution of E-M78 subhaplogroups was evaluated in a large sample of northern (Piedmont) and southern (Sicily) Italian males.

## Materials and methods

Sequence information for marker M78 and 9 Y-SNPs (M148, M224, V12, V13, V19, V22, V27, V32, and V65) representing final branches of the E1b1b1a haplogroup were drawn from the literature [15, 16, 18]; marker V36 was excluded since it was shown to be phylogenetically equivalent to V13 [16]. Primers for multiplex PCR and SBE reaction were designed with the Primer Express software (Applied Biosystems, Foster City, CA, USA). The main characteristics of PCR primers are described in Table S1 of Electronic supplementary material. In order to optimize multiplex PCR conditions, primers with homogeneous $T_m$ values (58.7±1.6°C) and expected amplicon size (between 96 and 141 bp) were chosen. Y-chromosomal specificity for each primer was first verified using the Basic Local Alignment Search Tool software (http://blast.ncbi.nlm.nih.gov). For Y-SNPs lying in regions of high homology with the X chromosome, PCR primers characterized by a mismatch at the 3′ end when compared with X-chromosomal sequences were preferentially selected. A further control was performed by singleplex amplification of each Y-SNP PCR primer pair using a female DNA template; absence of female DNA-derived PCR products was then checked by agarose gel electrophoresis. Primer concentrations in the multiplex PCR reaction are also shown in Table S1. Multiplex PCR amplification was performed in a final volume of 10 μl including 0.25 ng of template DNA, using the QIAGEN Multiplex PCR kit (QIAGEN, Hilden, Germany). The PCR protocol consisted of a 15-min pre-PCR heat step at 95°C, followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 58°C for 90 s, and extension at 72°C for 1 min, with a final 10-min extension step at 72°C. Of the PCR product, 1 μl was purified by incubation with 0.75 μl of *Escherichia coli* exonuclease I (ExoI; New England BioLabs, Ipswich, MA, USA) mix containing 5 U of ExoI in ExoI buffer 1× and 0.75 μl of calf intestinal alkaline phosphatase (CIP; New England BioLabs) mix containing 2.5 U of CIP in 1× NE buffer 3. The protocol consisted of 60 min at 37°C followed by enzyme inactivation at 80°C for 20 min.

The SBE reaction was performed using the SNaPshot Multiplex kit (Applied Biosystems) in a final volume of 5 μl containing 2.5 μl of purified multiplex PCR product and 0.2 μM of each SBE primer. The characteristics of SBE primers are shown in Table S2 of Electronic supplementary material. By addition at the 5′ end of each primer of a tail composed either of poly-T/poly-GCAT or a 40-bp-long neutral sequence [19], spatially separated DNA fragments ranging from 18 to 65 bases could be obtained. The SBE reaction consisted of 25 cycles of denaturation at 96°C for 10 s, annealing at 50°C for 5 s, and extension at 60°C for 10 s. SBE products were further purified by addition of 0.5 μl of CIP mix and incubation at 37°C for 60 min, followed by inactivation at 75°C for 15 min. Of the purified SBE product, 0.5 μl was run on an ABI Prism 310 Genetic Analyzer (Applied Biosystems) in a 47-cm capillary filled with POP-4 polymer using the E5 module. Electrophoresis parameters were set as follows: sample injection for 5 s at 15 kV and separation for 16 min at 15 kV and 60°C. Genescan-120 LIZ (Applied Biosystems) was used as internal size standard and data were analyzed by the Genescan Analysis software v. 3.7 (Applied Biosystems).

In order to evaluate the sensitivity of the assay, serial dilutions of male DNA at concentrations of 1 ng/μl and 500, 250, 100, and 50 pg/μl DNA were tested. Amplification and minisequencing of mixed samples containing ratios of female and male DNA ranging between 1:1 and 500:1 were also performed.

For the population study, the distribution of E-M78 and its subhaplogroups was analyzed in 326 unrelated Italian males originating from Sicily ($n$=236) and Piedmont ($n$=90). Y-STR loci included in the minimal haplotype (minHt) were also typed in the population

sample following the protocol described by Anslinger et al. [20].

## Results and discussion

Representative electropherograms obtained with the 10-plex Y-SNP assay in the Italian population sample are shown in Figure S3 of Electronic supplementary material. The observed SBE products were consistently longer than the expected size, as indicated in Table S3 of Electronic supplementary material. Such an anomaly in the mobility of short DNA fragments in capillary gels can be ascribed, as previously suggested by other authors, to the presence of a secondary structure in the extension products [21] and to the effect of fluorescent dyes conjugated to ddNTPs [22]. For the same reason, slight differences in size could be observed within single markers depending on the wild-type or mutated allelic status of the tested sample. M78 and M224 were the two markers showing the closest SBE peaks. However, due to the combination of the expected dyes associated to each extension product (G or A for M78 and T or C for M224, according to the ancestral or mutated status at the polymorphic site), the occurrence of ambiguous typing results could be avoided.

As also reported in Table S3, a reproducibility study conducted while typing the Italian population sample showed that for some Y-SNP markers the mean peak height of SBE products varied significantly between individuals carrying the wild-type and mutated allele. To overcome this problem, the concentration of multiplex PCR primers was arranged with the aim of maximizing the amplification of Y-SNPs showing potentially unfavorable polymorphic variants. Primer and PCR conditions were optimized in order to obtain balanced peak heights between different markers with an input of template DNA of about 0.25 ng. Variation of the initial DNA amount in the PCR reaction did not affect the balancing between amplification products, with the exception of V13, M78, and V65. As the quantity of template DNA increased, the first two Y-SNPs showed a relative increment of peak height/area compared to other markers, whereas the dimension of V65-specific peaks was progressively reduced, although constantly higher than 100 rfu when 1 ng of template DNA was used.

The sensitivity study indicated that complete SBE profiles with peaks >100 rfu could be obtained even in the presence of reduced amounts of template DNA (as little as 50 pg), comparable to DNA quantities extracted from forensic casework samples in LCN conditions [23]. Analysis of DNA admixtures demonstrated that full male Y-SNP profiles could be typed in mixed samples even in the presence of excess female DNA at a ratio of 500:1. The results of sensitivity and admixture studies are shown in Fig. S4 of Electronic supplementary material.

Haplogroup assignation by the minisequencing method in the Sicilian population subset was in concordance with Y-SNP typing results previously obtained in the same sample by denaturing high-performance liquid chromatography [24]. The frequencies of E-M78 (E1b1b1a) and its subhaplogroups in the Italian population sample are displayed in Table S4 of Electronic supplementary material. The overall occurrence of E1b1b1a in Piedmont (7.78%) and Sicily (11.44%) is in accordance with previous population studies [16, 25]. With regards to subhaplogroups, only E-V12*, E-V13*, E-V22*, and E-V65 were observed, the latter subhaplogroup exclusively in Sicily. Distribution of E-M78 subhaplogroups in the Sicilian sample closely resembles that previously reported by Cruciani et al. [16], whereas the presence of E-V12* in northern Italy (2.22%) is here described for the first time. The obtained results confirm the fact that even geographically restricted M78-derived variants, like E-V12* (observed at a frequency of 44.30% in southern Egypt), E-V22 (21.95% and 25% in Egypt and Ethiopia), and E-V65 (29.09% and 20% in Moroccan and Libyan Arabs, respectively), can be found at low though not negligible levels in distant populations [16]. Prediction of population of origin by Y-chromosomal haplogroup analysis in forensic casework should therefore be limited to intelligence purposes and initial suspect screening.

Since Y-STR haplotype diversity within subjects carrying a certain haplogroup is limited to mutations accumulated over generations in the founder haplotype [26], specific allele combinations may also be used to infer haplogroup status [27] and for prediction of geographic or ethnic origin of unknown stains [28]. Limitations to the use of Y-STR loci for haplogroup inference depend on the high mutation rate, which might homogenize haplotypes of different descent [29]. Combined analysis of Y-SNPs and Y-STRs in Italians carrying haplogroup E1b1b1a showed that, if considering minHt loci with the exclusion of biallelic DYS385, two cases of haplotype sharing between E-V12* and E-V13* individuals and one between E-V12* and E-V65 subjects were observed (Table S5 of Electronic supplementary material). These findings indicate that caution is needed when using STRs as surrogates of geographically informative SNPs unless only major clades of the Y phylogeny, generated by highly divergent founding haplotypes, are considered [30].

In conclusion, due to its high sensitivity, male specificity, and reduced amplicon size (<150 bp), the described SNP assay is suitable for forensic stain and degraded or ancient DNA analysis [31]. It can therefore be combined with available major Y-chromosomal haplogroup minisequencing systems, which already include E-M78 but not its derived

subhaplogroups [10, 12], in order to provide increased power of population discrimination with minimal DNA sample consumption.

**Acknowledgements** The continuing support of the Compagnia di San Paolo to C. Torre's laboratory is gratefully acknowledged.

## References

1. Prinz M, Boll K, Baum H, Shaler B (1997) Multiplexing of Y chromosome specific STRs and performance for mixed samples. Forensic Sci Int 85:209–218
2. Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4:598–612
3. Jobling MA (2001) Y-chromosomal SNP haplotype diversity in forensic analysis. Forensic Sci Int 118:158–162
4. Amigo J, Phillips C, Lareu M, Carracedo A (2008) The SNPforID browser: an online tool for query and display of frequency data from the SNPforID project. Int J Legal Med 122:435–440
5. Szibor R, Plate I, Schmitter H, Wittig H, Krause D (2006) Forensic mass screening using mtDNA. Int J Legal Med 120:372–376
6. Köhnemann S, Sibbing U, Pfeiffer H, Hohoff C (2008) A rapid mtDNA assay of 22 SNPs in one multiplex reaction increases the power of forensic testing in European Caucasians. Int J Legal Med 122:517–523
7. Dixon LA, Murray CM, Archer EJ, Dobbins AE, Koumi P, Gill P (2005) Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes. Forensic Sci Int 154:62–77
8. Sobrino B, Brión M, Carracedo A (2005) SNPs in forensic genetics: a review on SNP typing methodologies. Forensic Sci Int 154:181–194
9. Sanchez JJ, Børsting C, Hallenberg C, Buchard A, Hernandez A, Morling N (2003) Multiplex PCR and minisequencing of SNPs-a model with 35 Y chromosome SNPs. Forensic Sci Int 137:74–84
10. Brión M, Sanchez JJ, Balogh K et al (2005) Introduction of an single nucleodite polymorphism-based "Major Y-chromosome haplogroup typing kit" suitable for predicting the geographical origin of male lineages. Electrophoresis 26:4411–4420
11. Onofri V, Alessandrini F, Turchi C, Pesaresi M, Buscemi L, Tagliabracci A (2006) Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs. Forensic Sci Int 157:23–35
12. Wetton JH, Tsang KW, Khan H (2005) Inferring the population of origin of DNA evidence within the UK by allele-specific hybridization of Y-SNPs. Forensic Sci Int 152:45–53
13. King TE, Parkin EJ, Swinfield G et al (2007) Africans in Yorkshire? The deepest-rooting clade of the Y phylogeny within an English genealogy. Eur J Hum Genet 15:288–293
14. Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res 18:830–838
15. Cruciani F, La Fratta R, Torroni A, Underhill PA, Scozzari R (2006) Molecular dissection of the Y chromosome haplogroup E-M78 (E3b1a): a posteriori evaluation of a microsatellite-network-based approach through six new biallelic markers. Hum Mutat 27:831–832
16. Cruciani F, La Fratta R, Trombetta B et al (2007) Tracing past human male movements in northern/eastern Africa and western Eurasia: new clues from Y-chromosomal haplogroups E-M78 and J-M12. Mol Biol Evol 24:1300–1311
17. Cruciani F, La Fratta R, Santolamazza P et al (2004) Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. Am J Hum Genet 74:1014–1022
18. Underhill PA, Passarino G, Lin AA et al (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet 65:43–62
19. Lindblad-Toh K, Winchester E, Daly MJ et al (2000) Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. Nat Genet 24:381–386
20. Anslinger K, Keil W, Weichhold G, Eisenmenger W (2000) Y-chromosomal STR haplotypes in a population sample from Bavaria. Int J Legal Med 113:189–192
21. Konrad KD, Pentoney SL Jr (1993) Contribution of secondary structure to DNA mobility in capillary gels. Electrophoresis 14:502–508
22. Vallone PM, Just RS, Coble MD, Butler JM, Parsons TJ (2004) A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome. Int J Legal Med 118:147–157
23. Gill P, Whitaker J, Flaxman C, Brown N, Buckleton J (2000) An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. Forensic Sci Int 112:17–40
24. Di Gaetano C, Cerutti N, Crobu F et al (2008) Differential Greek and northern African migrations to Sicily are supported by genetic evidence from the Y chromosome. Eur J Hum Genet . doi:10.1038/ejhg.2008.120
25. Semino O, Magri C, Benuzzi G et al (2004) Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. Am J Hum Genet 74:1023–1034
26. Bosch E, Calafell F, Santos FR et al (1999) Variation in short tandem repeats is deeply structured by genetic background on the human Y chromosome. Am J Hum Genet 65:1623–1638
27. Forster P, Röhl A, Lünnemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B (2000) A short tandem repeat-based phylogeny for the human Y chromosome. Am J Hum Genet 67:182–196
28. Robino C, Crobu F, Di Gaetano C et al (2008) Analysis of Y-chromosomal SNP haplogroups and STR haplotypes in an Algerian population sample. Int J Leg Med 122:251–255
29. De Knijff P (2000) Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. Am J Hum Genet 67:1055–1061
30. Schlecht J, Kaplan ME, Barnard K, Karafet T, Hammer MF, Merchant NC (2008) Machine-learning approaches for classifying haplogroup from Y chromosome STR data. PLoS Comput Biol 4: e1000093
31. Bouakaze C, Keyser C, Amory S, Crubézy E, Ludes B (2007) First successful assay of Y-SNP typing by SnaPshot minisequencing on ancient DNA. Int J Legal Med 121:493–499