J Forensic Sci, May 2005, Vol. 50, No. 3
Paper ID JFS2004245
Available online at: www.astm.org

Elizabet Petkovski,[1,2] *Ing.; Christine Keyser-Tracqui,*[1] *Ph.D.; Rémi Hienne,*[2] *Ph.D.; and Bertrand Ludes,*[1] *M.D., Ph.D.*

SNPs and MALDI-TOF MS: Tools for DNA Typing in Forensic Paternity Testing and Anthropology*

ABSTRACT: DNA markers used for individual identification in forensic sciences are based on repeat sequences in nuclear DNA and the mitochondrial DNA hypervariable regions 1 and 2. An alternative to these markers is the use of single nucleotide polymorphisms (SNPs). These have a particular advantage in the analysis of degraded or poor samples, which are often all that is available in forensics or anthropology. In order to study the potential of SNP analysis in these fields, 41 SNPs were selected on the basis of following criteria: conservation, lack of phenotypic expression, and frequency of occurrence in populations. Thirty-six autosomal SNPs were used for genotyping 21 inclusionary and 3 exclusionary paternity cases. The behavior of 5 X-chromosome SNPs was analyzed in a French representative population. Our approach to SNP typing is a multiplex PCR based amplification followed by simultaneous detection by primer extension (PEX) analyzed by Matrix Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS). The selected autosomal SNPs showed independent inheritance and gave clear results in paternity investigation. All X-SNPs were useful as both paternity and identification markers. PEX and MALDI-TOF MS, with their high sensitivity, precision and speed, gave a powerful method for forensic and anthropological exploitation of biallelic markers.

KEYWORDS: forensic sciences, DNA typing, anthropology, single nucleotide polymorphisms, paternity testing, individual identification, MALDI-TOF MS

Various polymorphic markers on the human genome provide excellent means of discrimination between individuals. Among these are Variable Number of Tandem Repeats (VNTRs), autosomal and Y-chromosomal Short Tandem Repeats (STRs) (1–3), and the polymorphism of the variable mitochondrial D-loop nucleotide sequence (4,5). Despite the fact that single nucleotide variations (SNPs) are biallelic, they offer several attractive features as genetic markers. Due to the low mutation rate of SNPs, their information refers to longer periods of time compared to that obtained with STRs and VNTRs (6,7). This and their frequent occurrence in the human genome (8) make SNPs an attractive tool for parentage testing (9) and anthropology studies. In forensic paternity casework and anthropological investigations, where biological samples are often poor or degraded, the particular advantage of SNPs is that the studied DNA sequences are much shorter than those used for "classic" DNA analysis (10,11). Furthermore, it has been demonstrated that the information from 50 SNP arrays would be comparable to that obtained by existing STR multiplexes provided that they are chosen so that the allele frequencies range between 0.2 and 0.8 (12).

The aim of the present study is the evaluation of an SNP-panel for both forensic and anthropological studies. A panel of 36 autosomal and 5 X-chromosome informative biallelic markers was analyzed in order to establish individual specific profiles. SNPs were chosen to have the following characteristics: intronic position allied to stability against mutation and representation at high frequencies among populations. Guided by frequency analysis and physical maps provided by databases such as Applied Biosystems Assays-on-demand database and the SNP consortium (Table 1), we were able to make the assumption that they are inherited independently. The lack of linkage disequilibria between markers located on the same chromosome was confirmed by Genepop calculations using the obtained genotyping data.

Paternity studies and individual identifications require an analysis method that can provide information from an extended number of markers on each sample. If SNP typing is to be used in forensics and anthropology, it is essential that the investigations can be performed on small amounts (sub-nanogram) of DNA. If the polymerase chain reaction (PCR) is used, the amplification of all DNA fragments to be investigated must be done in very few amplification reactions. Our chosen detection and analysis technology was the Primer EXtension (PEX) method followed by Matrix Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS).

Materials and Methods

The analysis of selected SNPs included the following major steps: PCR amplification of the target SNP containing DNA fragment, which was to be shorter than 200bp; Double stranded amplicon purification by Genopure™ ds magnetic bead assisted technology (Bruker Daltonics); Allele specific primer extension (PEX); Single stranded product purification by Genopure™ oligo technology (Bruker Daltonics) and MALDI-TOF MS analysis.

For this study, the SNPs (Table 1) were chosen from either published results or from publicly available databases such as Applied Biosystems Assays-on-demand database and the SNP consortium. They have all been investigated and reported as not being linked by position or by function to genetic disease, and their statistics in different populations have been evaluated.

Thirty-six autosomal polymorphisms and 5 X-SNPs were initially selected for a sequencing test (Table 1), with the following

[1] Institute of Legal Medicine, University Louis Pasteur, 11 rue Humann, 67085 Strasbourg Cedex, France.
[2] CODGENE Laboratory, 11 rue Humann, 67085 Strasbourg Cedex, France.
* This work has been financially supported by the ALSACE region through a PRISME contract and a French national grant through a CIFRE contract.
Received 19 June 2004; and in revised form 2 Oct. 2004; accepted 16 Nov. 2004; published 6 April 2005.

TABLE 1—*Selected SNPs: chromosome localization and reported allelic frequencies.*

rs N°	Position	Alleles	Frequencies	rs N°	Position	Alleles	Frequencies
rs1020636	2p22.3	T/C	0.542/0.458	rs243	10q24.1	T/C	0.539/0.461
rs222	2q11.2	T/C	0.775/0.225	rs544021	11q13.1	C/G	0.567/0.433
rs227	3p26.1	G/A	0.780/0.220	rs882937	11q14.3	G/A	0.558/0.442
rs228	3p26.1	C/T	0.820/0.180	rs1793286	11q25	T/A	0.527/0.473
rs1551995	3p26.3	C/T	0.608/0.392	rs1801046	12p13	C/T	0.700/0.300
rs225	3q29	C/T	0.520/0.480	rs2123	12q14.1	A/G	0.330/0.670
rs230	4q34.1	G/A	0.608/0.392	rs811	12q21.33	A/G	0.653/0.347
rs724784	5q21.3	A/C	0.554/0.446	rs1924609	13q21.1	T/C	0.636/0.364
rs889012	5q31.1	A/C	0.517/0.483	rs6118	14q32.13	C/T	0.790/0.210
rs3792774	5q33.1	A/G	0.540/0.470	rs1053874	16p13.3	A/G	0.560/0.440
rs614570	6p12.2	T/C	0.400/0.600	rs820129	17q24.3	G/A	0.549/0.451
rs910170	6p21.2	C/T	0.522/0.478	rs276922	18q12.1	A/C	0.567/0.433
rs997556	7p12.1	T/C	0.618/0.382	rs1674139	19q13.4	C/T	0.619/0.381
rs2260	7q21.3	C/T	0.360/0.640	rs754	20q13.33	A/G	0.770/0.230
rs234	7q22.3	C/T	0.607/0.393	rs228047	21q22.3	C/T	0.753/0.247
rs725	7q34	A/G	0.729/0.271	rs618	Xq22.3	G/A	0.530/0.470
rs1039854	8q21.12	C/T	0.600/0.400	rs16282	Xq24	T/C	0.620/0.380
rs874746	8q24.3	C/A	0.392/0.608	rs616	Xq25	C/T	0.500/0.500
rs240	9q31.3	C/G	0.552/0.448	rs17379	Xq26.2	G/A	0.540/0.460
rs237	9q34.1	A/G	0.550/0.450	rs17407	Xq27.2	G/A	0.820/0.180
rs241	10q22.3	A/C	0.567/0.433				

requirements: to be silent (i.e., present on the DNA sequence but not expressed), to be located inside non-coding regions, to be at least 2 Mbp apart, to be as discriminatory as possible with allele frequencies ranging between 0.2 and 0.8, and to have complete sequence information on the regions flanking the target site available for PCR design. *In fine* the physical distances sharing loci on the same chromosome ranged between 2,3 Mbp and 226 Mbp.

We evaluated 5 SNPs located on the X chromosome with available information (Table 1) as markers for forensic and anthropology studies (13). The X-SNPs were selected from the ones reported as dimorphic with frequencies of about 50% (14).

DNA specimens collected from 21 unrelated French trios (father, mother and son) with independent evidence of paternity inclusion and 3 exclusion cases (unrelated individual, mother and son) were used for autosomal SNP analysis. The parental relationship between typed individuals had been previously proven or excluded by the currently used STR method. We studied X chromosome SNPs of 120 unrelated French men and 28 unrelated French women in order to evaluate their frequencies and information content in a French representative population.

DNA was purified from blood samples stored on FTA®-treated matrix (Whatman BioScience Ltd). PCR was performed in single reactions on 1.2 mm FTA® cardpunches after 3 washings with the manufacturer recommended FTA purification reagent, and 2 washings with TE light buffer (10 mM Tris-HCl pH 8 buffer and 0.1 mM EDTA).

Primers (18 to 22-base oligonucleotides) were designed according to the DNA sequence flanking the target SNP sites. Primers were designed so that the amplicons ranged in size from 50 to 200 base pairs with the OLIGO-4.0-s software (http://www.oligo.net/). Each target DNA fragment was tested in singleplex PCR and Big Dye Terminator™ reactions (Applied Biosystems) followed by capillary electrophoresis on an ABI 3100™ analyzer (Applied Biosystems) with primers outside those for final PCR and PEX reactions (200 to 300 bp fragments). This allowed us to verify the SNP nature, its flanking sequences and the primer hybridization region.

All primers were selected to have theoretical melting temperatures of 56 to 58°C, a purine:pyrimidine content close to 1:1 when possible, and no or low strength secondary structures ($\Delta G > -6$ kcal/mol). Primer pairs were tested for primer-dimer interactions, and their sequences were checked to avoid similarities with repetitive sequences or with other loci in the genome. HPLC purified primers for amplification were purchased from MWG Biotech.

DNA template on 1.2 mm FTA® cardpunch (Whatman BioScience) was amplified by PCR in a 30 μL reaction volume containing 1X Magnesium free reaction buffer; 2 mM $MgCl_2$; 200 μM each dNTP; 0.32 μM each amplification primer; 0.5 unit DNA Polymerase under the following conditions: enzyme activation for 15 min at 95°C followed by 35 cycles (95°C–30 sec; 51°C–45 sec; 72°C–30 sec) and 2 min at 72°C final elongation step. All PCR reactions as well as the analysis steps were repeated three times in order to confirm the reproducibility of the results. Following amplification, the products were observed under UV light after electrophoresis in a 3% standard agarose ethidium bromide containing gel in 1X TBE buffer (89 mM Tris base, 89 mM boric acid, 2 mM EDTA, pH 8.3).

The final setup of the multiplex PCR amplification included 1,2 mm DNA containing FTA® cardpunch (Whatman BioScience) in a 30 μL reaction volume containing 1X Magnesium free reaction buffer; 3 mM $MgCl_2$; 200 μM each dNTP; 0.2–0.4 μM each amplification primer; 1 unit DNA Polymerase. Each PCR was optimized to yield 6 to 10 equilibrated autosomal SNP containing fragments. Fragments containing SNPs on the X chromosome were studied in separate single PCR and PEX reactions.

All DNA amplifications were performed in a GeneAmp 9600 thermal cycler (Perkin-Elmer) using the following program: enzyme activation and denaturation for 15 min at 95°C followed by 35 cycles (95°C–30 sec; 51°C–45 sec; 72°C–30 sec) and 2 min at 72°C final elongation step. The primer concentrations in the multiplex reaction were adjusted in order to obtain equal amounts of each PCR product. The primer concentrations ranged from 0.2 to 0.4 μM.

PCR products were purified with magnetic bead DNA purification system Genopure™ ds (Bruker Daltonics) according to the manufacturer's recommendations and eluted in 10 μL elution buffer.

Primers for SNP detection were designed with the 3′ base corresponding to the last base before the possible polymorphism. The sequences of the primers were checked for the possibility of primer-dimer and hairpin formation and investigated in self-extension reactions (negative controls). In order to give better discrimination

TABLE 2—*Example of a triPEX design. 3 SNPs are to be detected by the 3 detection primers (rs228047, rs1674139 and rs234). In Table 2a are summarized SNPs, their 3′ flanking sequences and that of the dNTPs and ddNTPs necessary to the PEX reaction. In Table 2b are summarized potential products of the PEX reaction and their masses classified in ascending order. The column "code" summarizes the annotations used in Fig. 1 in order to simplify it.*

(2a)

SNP	Primer S + SNP + 3′ Flanking Region	Nucleotides in PEX Reaction
rs228047	5′-CGCACCCCGCAGACCT-3′-R-ttgga	dG, ddT, ddA
rs1674139	5′-GAGCTGACTTGCCGCC-3′-R-ggatt	dG, ddA
rs234	5′-ATGTGGCAGAGACTGAAT-3′-R-aaggg	dG, ddA

(2b)

Product/Allele	Product Sequence	Product Mass	Code
rs228047	CGCACCCCGCAGACCT	4772 Da	1
rs1674139	GAGCTGACTTGCCGCC	4858 Da	2
rs228047-A	CGCACCCCGCAGACCTddA	5085 Da	1a
rs1674139-A	GAGCTGACTTGCCGCCddA	5171 Da	2a
rs228047-G	CGCACCCCGCAGACCTdGddT	5406 Da	1b
rs234	ATGTGGCAGAGACTGAAT	5588 Da	3
rs234-A	ATGTGGCAGAGACTGAATddA	5901 Da	3a
rs1674139-G	GAGCTGACTTGCCGCCdGdGdGdd	6159 Da	2b
rs234-G	ATGTGGCAGAGACTGAATdGddA	6230 Da	3b

between the different PEX products we wished to increase the mass range of the extended primers. This was done by incorporating within the PEX media nonlabelled ddNTPs complementary to the nucleotide corresponding to one allele of the SNP and to the first 3′ position following the SNP, and nonlabelled dNTP complementary to the nucleotide corresponding to the other SNP allele.

Two μL of purified PCR product were used for the PEX reaction. Ten μL PEX medium was optimized for the extension of all target detection primers with 1X Magnesium free reaction buffer C (Solis BioDyne®); 10 mM $MgCl_2$(Solis BioDyne®); 200 μM appropriate dNTP and ddNTP; 1 μM each detection primer; 1 unit ThermiPol DNA Polymerase (Solis BioDyne®) in 35 cycles (95°C–10 sec; 50°C–30 sec; 72°C–30 sec) and 2 min at 72°C final elongation step.

The composition of the final medium varied according to the nature of the target SNPs, their 3′ flanking sequence and the number of SNPs to be detected per multiplexed reaction (cf example in Table 2). Each PEX was designed to detect simultaneously 3–4 SNPs with no interference from dNTPs and ddNTPs.

PEX products were desalted and purified with magnetic bead DNA purification system Genopure™ oligo (Bruker Daltonics) according to the manufacturers recommendations and eluted in 5 μL elution buffer.

One μL eluted product was deposited on an Anchor-Chip™ (384–400 μm) (Bruker Daltonics) with 1 μL matrix according to the dried droplet method. For analytes having mass greater than 3,5 kDa, we used 3-HPA (3-hydroxy-picolinic acid 10 g/L, di-ammonium-hydrogen-citrate 2 g/L) as matrix. Once dried at room temperature for 15 min, the samples were analyzed on an Ultraflex™ (Bruker Daltonics) instrument in linear positive mode. Calibration was performed with the primers present in negative control samples.

Divergence from Hardy-Weinberg expectations and linkage disequilibria across loci located on the same chromosome were tested thanks to a publicly available program based on a Markov chain method (15), the downloadable DOS version of the Genepop program (http://wbiomed.curtin.edu.au/genepop/). The Markov chain parameters were set to 1000 dememorizations, 1000 batches and 5000 iterations per batch for all tests, in order for the standard error (S.E.) of the estimates to be beneath 0,001.

In order to study the sensitivity of the SNPs typing method, we have tested the by now developed multiplex reactions on hair samples which yielded no complete STR profile.

Results

We found that the best target amplification results were obtained by increasing the concentration to 3 mM $MgCl_2$. Higher concentrations either inhibited the amplification or led to non-specific amplification (data not shown), while lower concentrations unbalanced the product yield. The results that we obtained were reproducible.

The magnetic bead DNA purification system Genopure™ ds was very convenient as it led to the absence of dNTP or primer excess, and thus prevented interference with downstream steps such as PEX or sequencing with Big Dye Terminator™ kit.

Figure 1 shows representative mass spectra of 3 autosomal SNPs in a paternity case. The 36 biallelic loci located on autosomes were successfully typed with minimal background signal. Heterozygous genotypes displayed peaks of roughly equal intensities at the expected time of flight points.

All the autosomal biallelic markers presented dimorphism and independent assortment (Tables 3 and 4). The calculated level of polymorphism of the studied population is 100%, the heterozygosity per locus ranged between 0.13 and 0.5 whereas the average heterozygosity of all loci was 0.46. No deviations from Hardy-Weinberg equilibrium were observed ($p > 0.05$) except for the locus rs754 (Table 3), which diverged significantly from expectations. Furthermore, the observed allelic frequencies for this locus did not correspond to the previously retained selection criteria.

Statistical linkage analysis suggested no linkage disequilibria ($p > 0.05$) among loci located on the same chromosome except for 2 pairs (rs2260/rs725 and rs18010/rs811). These 2 loci pairs showed a p value under 0.05 due to a population size bias, as children genotypes were not taken into account for these calculations, but presented no evidence of linkage disequilibria.

They showed no mutation through generations and confirmed the parentage relationships previously demonstrated by the STR method (cf example in Table 5). Reproducibility was tested, and consistent results were obtained by typing each SNP three times by MALDI-TOF MS for every DNA sample. In each experiment

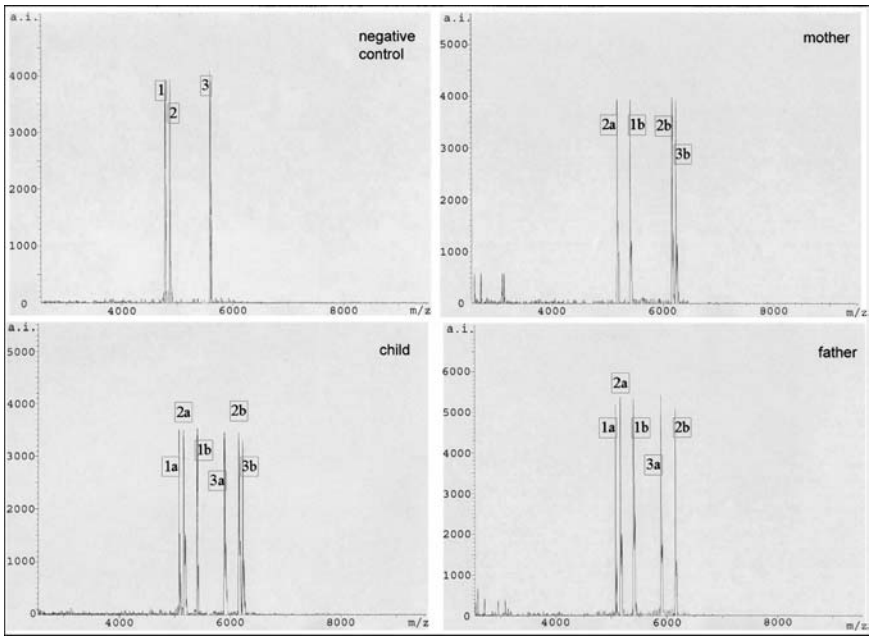

FIG. 1—*Example of spectra obtained by the analysis of 3 SNPs in a triPEX reaction. The peak annotations are summarized in Table 2b, column "code."*

a sample whose SNP profile had been determined by sequencing was included as a positive control and agreement of results was obtained in all cases.

The 5 X-Chromosome SNPs were all polymorphic with observed frequencies in concordance with selection criteria. Statistical linkage analysis (Table 4) suggested no linkage disequilibria ($p > 0.05$). Typing 120 hemizygous Frenchmen with these markers yielded 29 different profiles, whereas the 28 unrelated women were characterized by 23 different profiles due to dizygosity.

Typing hair samples on which no complete STR profile was obtained gave encouraging results as it allowed obtaining an 18 SNP profile (data not shown). The number of tested markers was limited by the multiplex format. Indeed, for the time being PCR reactions amplify 6 to 10 target SNP containing fragment. Enlarging the multiplex format will increase the success rate on difficult samples.

Discussion

In this study, we have evaluated the typing and the information content of a panel of SNPs by primer extension and Matrix Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry for forensic studies and potentially for anthropological investigations. The combination of single-point genomic markers detectable from minute amounts of DNA or degraded DNA, with specific multiplexed PEX detection and very sensitive, precise and rapid MALDI-TOF analysis brings a powerful

TABLE 3—*Observed allele frequencies, p-values for Hardy Weinberg equilibrium and heterozygosities for the 41 selected SNPs.*

Locus	Alleles	Frequencies	*p*-Value	Heterozygosity	Locus	Alleles	Frequencies	*p*-Value	Heterozygosity
rs1020636	C/T	0.438/0.562	0.2339	0.492312	rs243	C/T	0.480/0.520	0.4969	0.4992
rs222	C/T	0.276/0.724	0.2487	0.399648	rs544021	C/G	0.368/0.632	0.8053	0.465152
rs227	A/G	0.388/0.612	0.0525	0.474912	rs882937	A/G	0.414/0.586	0.6419	0.485208
rs228	C/T	0.842/0.158	0.1935	0.266072	rs1793286	A/T	0.414/0.586	1	0.485208
rs1551995	C/T	0.534/0.466	1	0.497688	rs1801046	C/T	0.697/0.303	0.7865	0.422382
rs225	C/G	0.605/0.395	0.475	0.47795	rs2123	A/G	0.289/0.711	0.7803	0.410958
rs230	A/G	0.474/0.526	0.8175	0.498648	rs811	A/G	0.730/0.270	1	0.3942
rs724784	A/C	0.546/0.454	0.4964	0.495768	rs1924609	C/T	0.309/0.691	0.5987	0.427038
rs889012	A/C	0.408/0.592	0.056	0.483072	rs6118	C/T	0.928/0.072	1	0.133632
rs3792774	A/G	0.322/0.678	0.2945	0.436632	rs1053874	A/G	0.336/0.664	0.2072	0.446208
rs614570	C/T	0.579/0.421	0.8157	0.487518	rs820129	A/G	0.684/0.316	0.5944	0.432288
rs910170	C/T	0.684/0.316	0.7949	0.432288	rs276922	A/C	0.500/0.500	0.6504	0.5
rs997556	C/T	0.354/0.646	0.4387	0.457368	rs1674139	C/T	0.559/0.441	0.49	0.493038
rs2260	C/T	0.441/0.559	0.642	0.493038	rs754	A/G	0.882/0.118	0.0074	0.208152
rs234	C/T	0.559/0.441	0.1059	0.493038	rs228047	C/T	0.533/0.467	0.1651	0.497822
rs725	A/G	0.632/0.368	1	0.465152	rs618	A/G	0.532/0.468	0.3422	0.497952
rs1039854	C/T	0.572/0.428	0.6409	0.489632	rs16282	C/T	0.235/0.765	0.1651	0.35955
rs874746	A/C	0.664/0.336	0.7927	0.446208	rs616	C/T	0.505/0.495	0.3458	0.49995
rs240	C/G	0.612/0.388	0.6296	0.474912	rs17379	A/G	0.548/0.452	0.3289	0.495392
rs237	A/G	0.579/0.421	0.3461	0.487518	rs17407	A/G	0.611/0.389	1	0.475358
rs241	A/C	0.467/0.533	0.3548	0.497822					
Average heterozygosity for autosomal loci				0.4458					
Average heterozygosity for X-loci				0.4656					

TABLE 4—*Statistical estimates of linkage disequilibria among SNP loci located on the same chromosome. The p-values lower than 0,05 are highlighted in gray.*

Chromosome	Loci	*p*-Value	Chromosome	Loci	*p*-Value
2	rs10206 & rs222	0.9850	9	rs240 & rs237	0.8296
3	rs15519 & rs225	0.8939	10	rs241& rs243	0.9243
3	rs227 & rs15519	0.3341	11	rs54402 & rs17932	0.9474
3	rs227 & rs225	0.9021	11	rs54402 & rs88293	0.3279
3	rs227 & rs228	0.3005	11	rs88293 & rs17932	0.3850
3	rs228 & rs15519	0.9252	12	rs18010 & rs2123	0.4699
3	rs228 & rs225	0.4208	12	rs18010 & rs811	0.0114
5	rs72478 & rs37927	0.0728	12	rs2123 & rs811	0.8445
5	rs72478 & rs88901	0.5350	X	rs618 & rs16282	0.55454
5	rs88901 & rs37927	0.6969	X	rs618 & rs616	0.07509
6	rs61457 & rs91017	0.2969	X	rs16282 & rs616	0.77089
7	rs2260 & rs234	0.1166	X	rs618 & rs17379	0.91078
7	rs2260 & rs725	0.0082	X	rs16282 & rs17379	0.44498
7	rs234 & rs725	0.5828	X	rs616 & rs17379	0.50707
7	rs99755 & rs2260	0.7576	X	rs618 & rs17407	0.18628
7	rs99755 & rs234	0.9806	X	rs16282 & rs17407	0.92234
7	rs99755 & rs725	0.8466	X	rs616 & rs17407	0.67171
8	rs10398 & rs87474	0.1328	X	rs17379 & rs17407	0.24060

TABLE 5—*Example of results obtained from the analysis of autosomal markers in 6 individuals belonging to 2 unrelated paternity cases. Case A is an inclusion and case B exclusion. Genotypes for only 18 SNPs are represented here, but data were obtained and analyzed for all selected markers. Of the 24 paternity trios tested, 3 were exclusionary showing at least 5 markers inconsistent with paternity or identity (highlighted in gray). Here, G stands for guanine, C for cytosine, A for adenine, T for thymine. Heterozygous genotypes are represented by the generic IUB codes, where M stands for A and C, Y for C and T, R for A and G and W for A and T. The probability of paternity (W) for the alleged father given the mother/son duo and is calculated by modifying the formulae used for STR analysis interpretation (29); DC stands for decimal code. Both values were obtained with results of 35 loci.*

	Mother A	Child A	Alleged Father A	Mother B	Child B	Alleged Father B
rs1020636	T	T	T	C	C	Y
rs222	C	C	C	Y	T	C
rs227	R	R	R	G	G	R
rs228	C	C	C	C	C	Y
rs1551995	T	T	Y	Y	C	T
rs230	R	A	A	R	R	R
rs889012	M	M	C	M	C	C
rs614570	C	C	Y	T	Y	C
rs910170	Y	Y	C	C	C	C
rs997556	Y	Y	T	T	Y	C
rs2260	Y	T	T	Y	T	T
rs234	C	Y	T	Y	Y	T
rs725	A	R	G	A	R	A
rs1039854	C	Y	Y	Y	Y	C
rs874746	M	A	A	M	M	M
rs240	S	C	S	G	G	C
rs237	R	A	A	A	R	G
rs241	C	C	M	M	A	A
TIC	29026382945785400	28957718903931700	29568694459502700	85234779193672400	50638423245972600	1327897181717 46000
W	Case A		99,9997149%	Case B		0%

method of DNA investigation for forensic and anthropological applications.

The selected autosomal markers gave clear results in paternity investigation except for rs754, which was excluded from downstream calculations (Table 5). All X-SNPs were shown to be useful paternity and identification markers.

In order to simplify and evaluate the results, the autosomal SNP data can be reduced to numerical information (16). By this method the following encoding scheme is used: genotype 0 (⇔ AA) corresponds to "homozygous with the allele of higher frequency," "homozygous with the allele of lower frequency" is designated as 1 (⇔ aa), and "heterozygous" as 2 (⇔ Aa). Thus genotyping a set of SNP markers produces a unique ternary identification code (TIC) for an individual, with each digit of the code representing the genotype at a specific marker. This ternary code can be easily transformed into a decimal number, which is simple to store (Table 5). Each decimal code is unique (except in the case of monozygotic twins) and allows a rapid and extremely simple genetic profile comparison. Our results corroborate the polymorphism of the 36 selected autosomal markers, their independent inheritance and their ability to provide individual specific genetic profiles. However, an enlargement of the DNA sample number would be necessary in order to assess the specificity of the profiles that we have obtained. Ternary encoding of genotype data, whose effectiveness for the straightforward identification of individuals is demonstrated here, should be compatible with any SNP analysis platform.

Numerical evaluation (Table 5) of the data was carried out using an adaptation of formulae applied to other investigation systems,

(17) together with the selected SNP frequencies observed in this study. In this paper we excluded the presence of null alleles, as all analysis reactions as well as sequencing fragments containing the amplification primer binding sites and the PEX analysis, repeated in triplicate, brought the same and concordant results. As in most genetic identification labs, we deemed that an incompatible result at one locus might occur as the result of a mutation, or due to the presence of a null allele, whereas more than 3 exclusionary loci (PI = 0) should be enough to prove a non-paternity (18). The individual PIs allow the calculation of the cumulative probability of paternity (W) by the same formulae used in other investigations (17). The calculated W's in the studied paternity cases all exceeded values of 99.99% with a panel of only 35 markers (the marker rs 754 presenting no Hardy Weinberg equilibrium and showing a lower allelic frequency less than 0.2 was excluded from these calculations) (Table 5). Increasing this panel to about 50 unlinked loci should provide W values comparable to the ones obtained by existing STR multiplexes (12).

Following the same simplification scheme as the one used for autosomal markers, the haplotypes of the X chromosomal SNPs can be translated into a binary identification code. Therefore each male individual would be characterized by an autosomal TIC and a sex-chromosomal BIC (X and Y chromosomal loci), whereas a woman could be designated by a unique TIC (incorporating both autosomal and X chromosomal loci). Stable and independent inheritance has also been observed in studies of X-SNPs, so that they can therefore be assessed as useful in genetic identification despite their sex-linked imbalance (13). This last feature may be an advantage, knowing that a daughter has to inherit her father's X-chromosome.

The above demonstrates the potential of X chromosome markers for solving some of the problems that may be encountered in forensic investigation. Technically, the implementation of X chromosome marker testing in forensic practice should not pose any insurmountable problems. However, although plentiful data are already available, the quantification of linkage disequilibria between X chromosome markers requires further intense research efforts.

Primer extension (19) for SNP analysis involves annealing a primer to a template PCR amplicon immediately downstream of an SNP position, and its extension by specifically incorporating nucleotides that are complementary to those contained in the PCR template immediately adjacent to the primer position. The major advantage of this technique is that the distinction between the variants is based on the high accuracy of the incorporated nucleotides rather than on differences in thermal stability between matched and mismatched probes. PEX assays give rise to allele-specific products that can be separated according to their intrinsic differences in mass without any labeling. This allows excellent discrimination between the homozygous and heterozygous genotypes and the assays are robust and tolerant of small variations in the reaction conditions. In other words the technique provides more robust signal-to-noise ratios when compared to other methods (20–22).

MALDI-TOF MS based approaches using PEX have been developed (22,23) in which extended primers are purified and detected by mass; the identity of the polymorph nucleotide is determined by measuring the mass of the extended primer, which is detected at a m/z value specific to the nucleotides added in the extension reaction. The majority of PEX products being monoprotonated, the observed value is virtually m. Minisequencing has become the most widely used MALDI-TOF MS based method for SNP analysis, and it has been successfully applied to the genotyping of SNPs located in biologically important genes (24,25). As the allele assignment is deduced from the intrinsic molecular mass of the products, the results are not influenced by external reaction conditions. The high levels of throughput that can be achieved for SNP genotyping in an automated MALDI-TOF MS system have been demonstrated (26).

One drawback of the original minisequencing approach is that the small mass differences between the ddNTPs (as little as 9 Da) that are incorporated in the extension reaction require pure extension products and high instrumental mass resolution to detect heterozygous genotypes accurately. Therefore we chose to apply a variation of the PEX reaction leading to a better discrimination between the extended primers: we incorporated nonlabelled ddNTPs complementary to the nucleotide corresponding to one allele of the SNP and to the first 3′ position following the SNP, and nonlabelled dNTP complementary to the nucleotide corresponding to the other SNP allele. This led to a time of flight separation not only according to the character of PEX products but also according to their lengths (27). This last technique avoids labeling and enzymatic shortening (28,29) of the products as it allows working in a higher range of masses, and the sensitivity necessary to distinguish between alleles is lowered. Primers of different lengths can be used in combination with mixtures of dNTP and ddNTPs designed to yield allele-specific primer extension products with clear differences in their molecular mass. This variant of the classical PEX reaction has also an advantage in specificity over other modified-PEX based methodologies. Indeed, it relies on the detection of both alleles, unlike allele dependent extension (30) where only one allele of a locus is physically detected per reaction, which implies genotype deduction or requires double reactions per locus corresponding to the 2 alleles.

The main disadvantages are the high prices of the instruments, and the "multi-step" procedures including purifications, as a limitation of MALDI-TOF is that PEX products must be rigorously desalted before measurement to avoid background.

The high throughput that is possible for SNP genotyping in an automated MALDI-TOF MS system has been demonstrated (31), and the major challenge in application of this method to forensic casework lies in multiplexing amplification and detection reactions. Although the method would allow analyzing a set of 50 SNPs per sample in very short time, generally forensic samples do not enclose enough template-DNA for 50 PCR reactions. Furthermore, the management of 50 PCRs followed by 50 PEXs par sample is very onerous, requiring significant expenditure of both consumables and time, and is error and contamination sensitive. The assay can be multiplexed if the PEX products for each SNP have nonoverlapping mass distributions.

The multiplex PCR SNP typing format presented here seems to be useful for forensic and anthropological investigations because small amounts of DNA and degraded DNA templates can be reliably typed. The chosen set of markers is not our final SNP set. Indeed, the SNP number has to be increased to 50 and the new markers have to be compatible with the already developed multiplexes regarding linkage disequilibria among markers and reaction setup.

These further selected markers must also be validated on a large sample set. Other necessary future steps development includes automation of detection and development of specific software for profile to sample attribution, adaptable to routine and high throughput forensic and anthropological work.

Note

In the manuscript above, 36 autosomal and 5X SNPs are evaluated as markers for forensic paternity testing and anthropology. Today, the studied panel has grown to 51 unlinked and informative autosomal SNPs analyzed by primer extension and MALDI-TOF MS, which raised the probabilities of paternity above 99,99999%

for the studied paternity trios. The completion of this final set included adding a sex dependent polymorphic locus and replacing markers that showed divergences from Hardy Weinberg equilibrium.

Acknowledgments

We gratefully thank the people who made it possible for us to run our investigations on their DNA by giving their informed consent prior to their inclusion in the study. We thank Dr. Van Dorsselaer for his assistance in our experiments in his lab at the CNRS (Ecole de Chimie des Polymères et des Matériaux).

References

1. Lee HC, Ladd C, Bourke MT, Pagliaro E, Tirnady F. DNA typing in forensic science. I. Theory and background. Am J Forensic Med Pathol 1994;15(4):269–82. [PubMed]
2. Jobling MA, Pandya A, Tyler-Smith C. The Y chromosome in forensic analysis and paternity testing. Int J Legal Med 1997;110(3):118–24. [PubMed]
3. Roewer L, Kayser M, de Knijff P, Anslinger K, Betz A, Caglia A, et al. A new method for the evaluation of matches in non-recombining genomes application to Y-chromosomal short tandem repeat haplotypes in European males. Forensic Sci Int 2000;114(1):31–43. [PubMed]
4. Lutz S, Weisser HJ, Heizmann J, Pollak S. mtDNA as a tool for identification of human remains. Identification using mtDNA. Int J Legal Med 1996;109(4):205–9. [PubMed]
5. Budowle B, Allard MW, Wilson MR, Chakraborty R. Forensics and mitochondrial DNA: applications, debates, and foundations. Annu Rev Genomics Hum Genet 2003;4:119–41. [PubMed]
6. Mountain JL, Knight A, Jobin M, Gignoux C, Miller A, Lin AA, et al. SNPSTRs: empirically derived, rapidly typed, autosomal haplotypes for inference of population history and mutational processes. Genome Res 2002;12:1766–72. [PubMed]
7. Brion M, Cao R, Salas A, Lareu MV, Carracedo A. New method to measure minisatellite variant repeat variation in population genetic studies. Am J Hum Biol 2002;14:421–4. [PubMed]
8. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 2001;409:928–33. [PubMed]
9. Reich DE, Schaffner SF, Daly MJ, McVean G, Mullikin JC, Higgins JM, et al. Human genome sequence variation and the influence of gene history, mutation and recombination. Nat Genet 2002;32:135–42. [PubMed]
10. Taillon-Miller P, Piernot EE, Kwok PY. Efficient approach to unique single-nucleotide polymorphism discovery. Genome Res 1999;9(5):499–505. [PubMed]
11. Syvanen AC, Sajantila A, Lukka M. Identification of individuals by analysis of biallelic DNA markers, using PCR and solid-phase minisequencing. Am J Hum Genet 1993;52:46–59. [PubMed]
12. Gill P. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. Int J Legal Med 2001;114(4–5):204–10. [PubMed]
13. Szibor R, Krawczak M, Hering S, Edelmann J, Kuhlisch E, Krause D. Use of X-linked markers for forensic purposes. Int J Legal Med 2003;117(2):67–74. [PubMed]
14. Taillon-Miller P, Kwok PY. A high-density single-nucleotide polymorphism map of Xq25–q28. Genomics 2000;65(3):195–202. [PubMed]
15. Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 1992 Jun;48(2):361–72. [PubMed]
16. Griffin TJ, Smith LM. Genetic identification by mass spectrometric analysis of single-nucleotide polymorphisms ternary encoding of genotypes. Anal Chem 2000;72(14):3298–302. [PubMed]
17. White LD, Shumaker JM, Tollett JJ, Staub RW. Human identification by genotyping single nucleotide polymorphisms (SNPs) using an apex microarray. Proceedings of the Ninth International Symposium on Human Identification, October 8–10, 1998. Profiles in DNA 1999; 203.
18. Brenner C. http://dna-view.com/
19. Griffin TJ, Smith LM. Single-nucleotide polymorphism analysis by MALDI-TOF mass spectrometry. Trends Biotechnol 2000;18(2):77–84. [PubMed]
20. Lareu M, Puente J, Sobrino B, Quintans B, Brion M, Carracedo A. The use of the LightCycler for the detection of Y chromosome SNPs. Forensic Sci Int 2001;118(2–3):163–8. [PubMed]
21. Wasson J, Skolnick G, Love-Gregory L, Permutt MA. Assessing allele frequencies of single nucleotide polymorphisms in DNA pools by pyrosequencing technology. Biotechniques 2002;32(5):1144–6, 1148, 1150.
22. Haff LA, Smirnov IP. Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. Genome Res 1997;7:378–88. [PubMed]
23. Higgins GS, Little DP, Koster H. Competitive oligonucleotide single-base extension combined with mass spectrometric detection for mutation screening. Biotechniques 1997;23:710–4. [PubMed]
24. Little DP, Braun A, O'Donnell MJ, Koster H. Mass spectrometry from miniaturized arrays for full comparative DNA analysis. Nat Med 1997;3:1413–16. [PubMed]
25. Little DP, Braun A, Darnhofer-Demar B, Frilling A, Li Y, McIver RT, Jr, et al. Detection of RET proto-oncogene codon 634 mutations using mass spectrometry. J Mol Med 1997;75:745–50. [PubMed]
26. Bray MS, Boerwinkle E, Doris PA. High-throughput multiplex SNP genotyping with MALDI-TOF mass spectrometry: practice, problems and promise. Hum Mutat 2001;17(4):296–304. [PubMed]
27. Pusch W, Kraeuter KO, Froehlich T, Stalgies Y, Kostrzewa M. Genotools SNP manager: a new software for automated high-throughput MALDI-TOF mass spectrometry SNP genotyping. Biotechniques 2001;30(1):210–5. [PubMed]
28. Fei Z, Ono T, Smith LM. MALDI-TOF mass spectrometric typing of single nucleotide polymorphisms with mass-tagged ddNTPs. Nucleic Acids Res 1998;26:2827–8. [PubMed]
29. Sauer S, Gut IG. Genotyping single-nucleotide polymorphisms by matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci 2002;782(1–2):73–87. [PubMed]
30. Blondal T, Waage BG, Smarason SV, Jonsson F, Fjalldal SB, Stefansson K, et al. A novel MALDI-TOF based methodology for genotyping single nucleotide polymorphisms. Nucleic Acids Res 2003;31(24):e155. [PubMed]
31. Maksymowych WP, Reeve JP, Reveille JD, Akey JM, Buenviaje H, O'Brien L, et al. High-throughput SNP analysis of the IL1RN locus in patients with ankylosing spondylitis by matrix-assisted laser desorption ionization-time-of-flight mass spectrometry. Arthritis Rheum 2003;48(7):2011–8. [PubMed]

Additional information and reprint requests:
Elizabet Petkovski, Ing.
Institute of Legal Medicine, University Louis Pasteur
11 rue Humann, 67085 Strasbourg Cedex
France
E-mail: petkovski@ulp.u-strasbg.fr