

# Development of SNP-based human identification system

Jae-Jung Kim · Bok-Ghee Han · Hae-In Lee ·  
Han-Wook Yoo · Jong-Keuk Lee

Received: 3 August 2009 / Accepted: 21 October 2009 / Published online: 18 November 2009  
© Springer-Verlag 2009

**Abstract** Single nucleotide polymorphisms (SNPs) appeal to the forensic DNA community because of their abundance in the human genome, low mutation rate, small amplicon size, and feasibility of high-throughput genotyping technologies. In an initial screening, we identified six SNP markers of sex determination by resequencing the amelogenin genes and the zinc finger protein genes located on the sex chromosomes. Furthermore, for use in human identification, we selected 30 highly polymorphic autosomal SNP markers from among a human population and examined the potential utility of these SNP markers for human identification. The combined mean match probability of 30 SNP markers was  $4.83 \times 10^{-13}$ . Using genotyping data from 8,842 unrelated Korean individuals, we also found that discrimination power increased 10-fold for the addition of every five SNP markers in human identification. In this

study, we demonstrated that SNP markers are very useful for sex determination and human identification, even in a very homogeneous population.

**Keywords** Single nucleotide polymorphism · Human identification · Sex determination · Korean

## Introduction

Genetic tests of human identification are critical to the field of forensic science. Currently, short tandem repeat (STR) markers are routinely used. These markers are highly informative because of the large number of alleles within various populations [1]. However, they are limited by their high mutation rate, difficulty with regards to multiplexing, and the need for large amplification products, which limits the use of degraded samples. Recently, single nucleotide polymorphisms (SNPs) have been promoted as useful genetic markers for human identification. SNPs have additional applications such as the determination of the geographic origin of a sample through the use of ancestry-informative marker SNPs [2]. Single nucleotide polymorphisms have low mutation rates and rely on short amplicons (i.e., less than 100 bp), which allow for the use of degraded DNA samples and high-throughput genotyping technologies [3–6]. To date, several studies have described preliminary SNP panels, containing autosomal and mitochondrial SNP markers as many as 52 markers, for use in human identification and parentage testing [7–14]. Furthermore, selecting a SNP with high heterozygosity and low differences in allele frequency for all populations (i.e., low  $F_{st}$ ) is important for the development of an efficient SNP marker system with high discriminating power [15]. In the

---

Jae-Jung Kim and Bok-Ghee Han contributed equally to this work.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00414-009-0389-9) contains supplementary material, which is available to authorized users.

---

J.-J. Kim · H.-I. Lee · J.-K. Lee (✉)  
Asan Institute for Life Sciences,  
University of Ulsan College of Medicine,  
388-1 Pungnap-2-Dong, Songpa-Gu,  
Seoul 138-736, Korea  
e-mail: cookie\_jklee@hotmail.com

B.-G. Han  
Center for Genome Science, National Institute of Health,  
Seoul, Korea

H.-W. Yoo  
Department of Pediatrics,  
University of Ulsan College of Medicine,  
Seoul, Korea

present study, we tested a set of SNP markers for human identification in a homogenous Korean population.

## Materials and methods

### Genomic DNA samples and large-scale SNP genotype data

To identify SNP markers of sex, we used 12 DNA samples from CEPH families (GM06994, GM07022, GM07347, GM07357, GM12962, GM12960, GM07000, GM07056, GM07346, GM07345, GM12863, and GM12861) obtained from the Coriell Institute for Medical Research (<http://www.coriell.org/>) and 12 DNA samples from unrelated Korean individuals obtained from the Biobank for Health Sciences at the Center for Genome Sciences in Seoul, Korea. To validate the selected SNP markers' suitability for sex determination and human identification, we obtained

both unrelated Korean DNA samples ( $n=960$ ) and Affymetrix Genome-Wide Human SNP Array 5.0 data from 8,842 Korean individuals [16] from the Center for Genome Sciences in Seoul, Korea. All DNA samples used in this study were isolated from blood samples or Epstein-Barr virus-transformed B-lymphocytes using standard DNA extraction methods. The study protocol was approved by the institutional review board of the Asan Medical Center.

### Resequencing

Sequencing analysis of the genes encoding amelogenin (i.e., AMELX and AMELY) and the zinc finger proteins (i.e., ZFX and ZFY) was performed using genomic sequences from GenBank (<http://www.ncbi.nlm.nih.gov/>). PCR primers specific to highly homologous target regions were designed using Primer3 software [17] (<http://frodo.wi.mit.edu/>). The PCR reactions were performed in a 20  $\mu$

**Fig. 1** Alignment of the genomic DNA sequences of the amelogenin genes (i.e., AMELX and AMELY) (a) and the zinc finger protein genes (i.e., ZFX and ZFY) (b). *Plus sign* indicate identity between two sequences. *Underlined sequences* indicate primer positions. *Large bold sequences* indicate polymorphic sites within and between AMELX and AMELY, and ZFX and ZFY loci, as identified in a sequencing analysis of 24 individuals

<b>a</b>	
AMELX	ACTGCTGCTTCTCTGGTTGGAGTCACCTGAGCCAAATGGTAAACCTGCCTCTCTGTTTCTC
AMELY	ACTGCTGCTTCTCTGGTTGGAGTCACCTGAGCCAAATGGTAAACCTGCATCTCTGTTTCTC
	*****
AMELX	ACCAGTAC <b>C</b> CTTCTCTATGGTTACGAGCCCATGGGTGGATGGCTGCACCACCAATCATCC
AMELY	ACCAGTAC <b>T</b> CTTCTCTATGGTTACGAGCCCATGGGTGGATGGCTGCACCACCAATCATCC
	*****
	<b>AMEL-1</b>
AMELX	CCGTG <b>C</b> TGTCCCAACAGCACCCCGACTCACACCCTGCAG <b>C</b> CTCATCACCACATCCCAG
AMELY	CCGTG <b>G</b> TGTCCCAACAGCACCCCTGACTCACACCCTGCAG <b>T</b> CTCATCACCACATCCCAG
	*****
	<b>AMEL-2</b> <span style="float: right;"><b>AMEL-3</b></span>
AMELX	<u>TGGTGCCAGCTCAGCAGCCC</u>
AMELY	<u>TGGTGCCAGCTCAGCAGCCC</u>
	*****
<b>b</b>	
ZFX	ACCAACAAGAAGATAAGTTTACACAACCACCTGGAGAGCCACAAGCTGACCAGCAAGGCA
ZFY	ACCAATAAGAAGATAAGTTTACATAACCACCTGGAGAGCCACAAGCTGACCAGCAAGGCA
	*****
ZFX	GAGAAGGCCATTGAATGCGATGAGTGTGGGAAGCATT <b>T</b> CTCTCATGCAGGGCTTTGTTT
ZFY	GAGAAGGCCATTGAATGTGATGAGTGTGGGAAGCATT <b>T</b> TCTCATGCAGGGCTTTGTTT
	*****
	<b>ZF-1</b>
ZFX	ACTCACAAAATGGTGATAAGGAAAAAGGAGCCAAACAAAATGCACAAAGTGTAAATTCTGT
ZFY	ACTCACAAAATGGTGATAAGGAAAAAGGGCCAAACAAAATGCACAAAGTGTAAATTCTGT
	*****
ZFX	GAATA <b>C</b> GAGACAGCTGAACA <b>A</b> GGGTTATTGAATCGCCACCTCTTGGCAGT <b>C</b> CCACAGCAAG
ZFY	GAATA <b>T</b> GAGACAGCTGAACA <b>G</b> GGGTTATTGAATCGCCACCTCTTGGCAGT <b>C</b> CCACAGCAAG
	*****
	<b>ZF-2</b> <span style="float: right;"><b>ZF-3</b></span>
ZFX	<u>AACTTCCTCATATTGTTGTGGAGTGTGGTAA</u>
ZFY	<u>AACTTCCTCATATTGTTGTGGAGTGTGGTAA</u>
	*****

volume containing 20 ng genomic DNA, 0.15  $\mu\text{M}$  of each primer, 100  $\mu\text{M}$  of each dNTP, 1 unit of AmpliTaq Gold™ (Applied Biosystems, Foster City, CA, USA), and reaction buffer containing 1.5 mM  $\text{MgCl}_2$ . The reactions were performed at 94°C for 5 min, followed by 35 cycles of 94°C for 45 s, 55°C for 30 s, and 72°C for 1 min. The reactions concluded with a final extension step at 72°C for 10 min. The PCR products (1  $\mu\text{l}$ ) were directly used as sequencing templates using the BigDye® terminator v3.1 cycle sequencing kit (Applied Biosystems, Foster City, CA, USA) on an ABI3730 DNA sequencer (Applied Biosystems, Foster City, CA, USA), according to the manufacturer's instructions. DNA polymorphisms were identified using the PolyPhred program (<http://droog.gs.washington.edu/PolyPhred.html>) after sequence chromatograms were base-called using the Phred program and assembled using the Phrap program [18].

#### Selection of highly informative candidate SNPs for human identification

To select candidate SNP markers for human identification, we initially used an Affymetrix 500 K SNP dataset that contained genotyping information from 2,000 Korean individuals and allele frequencies from four HapMap populations (Japanese, Han Chinese, African, and European). The following SNP selection criteria were used to choose candidate loci: (1) minor allele frequency (MAF) > 0.45 in all ethnic groups, (2) heterozygosity > 0.45, (3) Hardy–Weinberg equilibrium (HWE)  $p$  value > 0.1, (4) genotype call rate = 100%, and (5) physical distance between SNP markers > 50 Mb. Based on these criteria, we selected 30 SNPs for validation in a new sample set ( $n=960$ ).

#### SNP genotyping

Genotyping of selected SNP markers for sex determination and human identification in 960 individuals was performed using the custom-designed Illumina VeraCode GoldenGate Assays (Illumina Inc, San Diego, CA, USA) [19], according to the manufacturer's instructions. Genotype clustering and calling were performed using BeadStudio software (Illumina Inc, San Diego, CA, USA).

#### Statistical analysis

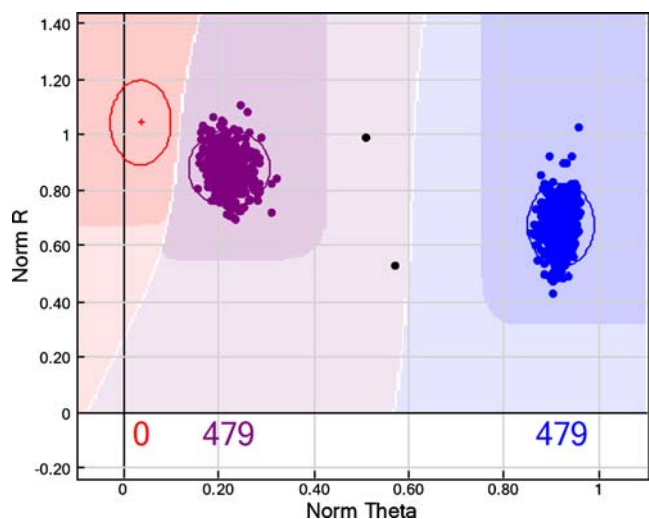
Deviation from Hardy–Weinberg expectation was determined using  $\chi^2$  tests. The match probability that two unrelated individuals will have the same multi-locus genotype was calculated as previously described [15]. Statistical analysis was performed using SPSS programs (version 14.0; SPSS Inc., Chicago, IL, USA).

## Results

### Identification of SNPs for sex determination

To identify SNP markers for sex determination, we initially selected genes encoding amelogenin (i.e., AMELX and AMELY) and the zinc finger proteins (i.e., ZFX and ZFY). Highly homologous target regions of the selected genes were chosen by multiple alignments (Fig. 1). PCR primers were designed to amplify and sequence approximately 150 to 250 bp of the target regions encompassing single nucleotide differences in genes on the X and Y chromosomes. Through the sequence comparison of 24 individuals, including 12 females and 12 males, we identified three bases each on the genes encoding amelogenin (i.e., AMEL-1, AMEL-2, and AMEL-3) and the zinc finger proteins (i.e., ZF-1, ZF-2, and ZF-3) that distinguished males from females (Supplementary Figs. 1 and 2).

To evaluate whether the identified markers could be used to determine sex, we tested four (i.e., AMEL-1, AMEL-2, ZF-1 and ZF-2) of six candidate markers using DNA samples of 960 individuals. Three markers (i.e., AMEL-1, AMEL-2 and ZF-1) yielded successful genotyping results, while one marker (i.e., ZF-2) failed. The genotyping results for the AMEL-1, AMEL-2 and ZF-1 markers matched the expected sex genotypes (male = heterozygote, female = homozygote) in the 960 individuals tested (Fig. 2). Thus, these validated markers can be used for sex determination in the SNP-based genetic test.



**Fig. 2** Genotyping of the AMEL-2 marker in 960 individuals using the Illumina VeraCode GoldenGate Assay. Raw data for the AMEL-2 marker are shown. Purple and blue dots indicate heterozygotes and homozygotes, respectively. Black dots indicate genotyping failure. All genotyped individuals yielded results corresponding to their recorded sex. The AMEL-1 and ZF-1 markers exhibited the same patterns of genotype clusters (data not shown)

**Table 1** Human identification by 30 highly polymorphic SNP markers

SNP (rs#)	Chr.	Position	Allele_1/2	Korean (n=2,000)		HWE_P	Match probability	Chinese		Japanese		European		African	
				Freq_I	Heterozygosity			Freq_I	Freq_I	Freq_I	Freq_I	Freq_I	Freq_I		
rs7532151	1	89161532	A/C	0.492	0.500	0.868	0.375	0.489	0.478	0.470	0.510	0.470	0.510	0.470	0.510
rs4846468	1	216462550	C/T	0.502	0.500	0.857	0.375	0.523	0.544	0.510	0.540	0.510	0.540	0.510	0.540
rs6751657	2	33258655	C/T	0.508	0.500	0.377	0.375	0.465	0.512	0.511	0.530	0.511	0.530	0.511	0.530
rs10185531	2	106105267	C/T	0.479	0.499	0.678	0.375	0.533	0.456	0.470	0.500	0.470	0.500	0.470	0.500
rs7652776	3	2716024	C/G	0.467	0.498	0.800	0.376	0.478	0.533	0.540	0.490	0.540	0.490	0.490	0.490
rs6443222 <sup>a</sup>	3	9138267	A/G	0.481	0.499	0.704	0.375	0.477	0.476	0.479	0.479	0.479	0.479	0.479	0.479
rs17497475	4	18851337	A/C	0.455	0.496	0.932	0.377	0.489	0.522	0.530	0.480	0.530	0.480	0.480	0.480
rs1350191	4	155026774	A/G	0.486	0.500	0.991	0.375	0.467	0.455	0.479	0.489	0.479	0.489	0.479	0.489
rs2565007	5	53856024	A/C	0.458	0.496	0.557	0.377	0.533	0.456	0.490	0.470	0.490	0.470	0.470	0.470
rs4607417	6	42086252	C/T	0.531	0.498	0.412	0.376	0.533	0.456	0.510	0.500	0.510	0.500	0.510	0.500
rs1790006	6	162297818	A/G	0.470	0.498	0.204	0.376	0.478	0.511	0.500	0.540	0.500	0.540	0.500	0.540
rs2813838	7	24115817	C/G	0.457	0.496	0.514	0.377	0.511	0.456	0.500	0.531	0.500	0.531	0.500	0.531
rs2267708	7	124179748	C/T	0.511	0.500	0.607	0.375	0.489	0.533	0.550	0.520	0.550	0.520	0.520	0.520
rs1293288	8	11755937	A/G	0.482	0.499	0.611	0.375	0.467	0.489	0.530	0.450	0.530	0.450	0.450	0.450
rs7849782	9	103467085	C/G	0.466	0.498	0.803	0.376	0.489	0.456	0.470	0.540	0.470	0.540	0.470	0.540
rs7907658	10	92782368	A/C	0.524	0.499	0.552	0.376	0.511	0.544	0.510	0.520	0.510	0.520	0.510	0.520
rs550840	11	63958215	C/T	0.473	0.499	0.999	0.376	0.500	0.500	0.490	0.520	0.490	0.520	0.490	0.520
rs543840	11	115269696	C/T	0.525	0.499	0.271	0.376	0.547	0.500	0.480	0.540	0.480	0.540	0.480	0.540
rs734075	12	4367308	A/C	0.469	0.498	0.585	0.376	0.467	0.467	0.521	0.490	0.521	0.490	0.490	0.490
rs1151849	12	119824982	A/G	0.454	0.496	0.352	0.377	0.477	0.478	0.480	0.500	0.480	0.500	0.480	0.500
rs7328030	13	111067445	A/C	0.517	0.499	0.248	0.375	0.533	0.533	0.540	0.450	0.540	0.450	0.450	0.450
rs978511	14	61773670	C/G	0.455	0.496	0.463	0.377	0.456	0.511	0.470	0.520	0.470	0.520	0.470	0.520
rs7164801	15	22608408	G/T	0.503	0.500	0.324	0.375	0.500	0.489	0.460	0.480	0.460	0.480	0.460	0.480
rs955665	15	77603212	C/T	0.535	0.498	0.303	0.376	0.500	0.533	0.510	0.542	0.510	0.542	0.510	0.542
rs4791495	17	12130882	A/G	0.502	0.500	0.532	0.375	0.467	0.533	0.450	0.510	0.450	0.510	0.450	0.510
rs4647887	17	72070401	A/G	0.532	0.498	0.245	0.376	0.477	0.544	0.550	0.550	0.550	0.550	0.550	0.550
rs1785745	18	21671131	A/G	0.486	0.500	0.993	0.375	0.542	0.465	0.543	0.511	0.543	0.511	0.511	0.511
rs7230112	18	71610565	C/T	0.535	0.498	0.467	0.376	0.500	0.534	0.480	0.510	0.480	0.510	0.480	0.510
rs8113496	19	34512369	A/G	0.482	0.499	0.919	0.375	0.467	0.489	0.540	0.467	0.540	0.467	0.467	0.467
rs2327088	20	8714882	A/G	0.491	0.500	0.940	0.375	0.522	0.522	0.450	0.480	0.450	0.480	0.450	0.480
CPI (combined power of identity)							1.76828E-13								
CPD (combined power of discrimination)							0.9999999999982300000 (1 in 5.65522E+12)								

The frequencies of other ethnic groups (i.e., Chinese, Japanese, Europeans, and Africans) were obtained from HapMap data (<http://www.hapmap.org>)

*HWE* Hardy–Weinberg equilibrium

<sup>a</sup> rs6443222 failed our SNP genotyping validation using the Illumina VeraCode GoldenGate Assay

### Selection of 30 highly informative SNP markers for human identification

To select highly informative SNP markers for human identification, we screened Affymetrix 500 K SNP genotype data that included genotyping results from 2,000 Korean individuals and allele frequencies from four HapMap populations (i.e., Japanese, Han Chinese, African, and European) and identified 56 highly polymorphic SNP markers that met the criteria described in the “Materials and methods”. Of these, 30 autosomal SNP markers were selected for validation in a new sample set ( $n=960$ ; Table 1). These markers were distributed across 19 different chromosomes containing approximately one to two SNPs per chromosome.

### Validation of the selected 30 SNP markers for human identification

The 30 selected SNP markers were tested in two independent Korean sample sets (i.e., 960 individuals and 8,842 individuals) to estimate their potential for human identification. The set of 960 individuals was genotyped using the custom-designed Illumina VeraCode GoldenGate Assay. Among the 30 SNP markers, 29 yielded successful genotyping results but one marker (rs6443222) failed. The allele and genotype frequencies of each SNP locus were almost same as those shown in Table 1 (shown in Supplementary Table 1). As expected, all SNP markers exhibited even allelic distributions. The combined mean match probability of 29 SNPs was  $4.83 \times 10^{-13}$ , corresponding with a combined power of discrimination of 0.99999999999517. When we compared the 960 individuals using 15 SNPs, only one pair (two individuals) had identical genotypes, indicating that these 15 SNP markers can discriminate among 1,000 individuals (data not shown).

The second sample set, composed of 8,842 individuals, was previously genotyped using the Affymetrix Genome-Wide Human SNP Array 5.0 [16]. As we had already selected candidate SNPs for human identification from the Affymetrix SNP Array 5.0 Genotype data, we used the same SNP markers and genotype data for human identification analysis. The combined mean match probability for the 30 selected SNPs was  $1.78 \times 10^{-13}$  and the combined power of discrimination was 0.99999999999822 (Supplementary Table 2). Furthermore, as shown in Table 2, each additional five SNP markers increased the discrimination power of SNP-based human identification by 10-fold (i.e., 10 and 15 SNP markers discriminated approximately 100 and 1,000 individuals, respectively). When 20 SNP markers were used, all 8,842 of the individuals tested were completely distinguished.

**Table 2** SNP-based human identification by 30 highly polymorphic SNP markers in 8,842 unrelated Korean individuals

No. of SNPs	No. of testing samples ( $n$ )	No. of samples showing identical genotypes <sup>a</sup>	Average identical (%)
10	100	0–0–2	0.67%
	1,000	54–38–39	4.37%
	8,842	2,769	31.32%
15	100	0–0–0	0.00%
	1,000	0–0–2	0.07%
	8,842	20	0.23%
20	100	0–0–0	0.00%
	1,000	0–0–0	0.00%
	8,842	0	0.00%

<sup>a</sup> We randomly selected three sets of samples that consisted of either 100 or 1,000 samples (total  $n=8,842$ )

### Discussion

Genetic tests for human identification are critical to the fields of forensic science and paternity testing [3, 12]. In addition, DNA-based human identification is necessary for efficient sample identification in biobanks that handle large numbers of biological samples [20, 21]. In this study, we tested the power of several SNP markers for sex determination and human identification in a Korean population. Our findings revealed that sex-specific and highly polymorphic SNP markers can discriminate individuals as efficient as STR markers when we used more SNP markers.

The genes encoding amelogenin and zinc finger proteins are routinely used for sex determination [22–24]. For STR-based determination of sex, 6 bp deletion polymorphisms located in intron 3 of the amelogenin gene, on the X chromosome, can generally be detected by size differences [23]. Previous studies have performed SNP-based sex determination by examining differences in the sequences of the amelogenin genes located on the X and Y chromosomes [9]. But the same DNA site was not perfectly matching with sex information in our 24 testing samples (C/T genotype in 12 female samples=11–1–0 and C/T genotype in 12 male samples=0–12–0). To identify optimal SNP markers for sex determination, we sequenced highly homologous regions of the genes encoding amelogenin (i.e., AMEL-X and AMEL-Y) and the zinc finger proteins (i.e., ZFX and ZFY). After resequencing, a total of six DNA markers were selected as potential SNP markers for sex determination. Those SNP markers were further tested in a large sample set ( $n=960$ ) and were able to perfectly match sex information. Sex-determining SNP markers will prove quite useful for the development of a SNP-based genetic test.

Although SNP markers are less informative than STR markers, SNP markers have many advantages for human



identification, including low mutation rate, fast genotyping, easy multiplexing, and small amplification size, which allows for the use of degraded DNA samples [3–6]. Furthermore, newly identified SNP markers for several phenotypic traits (e.g., blood type, height and skin color) can be easily integrated into the SNP-based human identification system as demonstrated in previous studies [9, 25]. The relative low power of SNP markers to discriminate among individuals can be easily overcome by multiplexing several (i.e., 40 to 50) SNP markers using the MassArray system (Sequenom), the SNPlex system (ABI), or a similar method. Our findings revealed that highly polymorphic 40 SNP markers have a combined power of identity (CPI) value of  $9.14 \times 10^{-18}$ , comparable with that of a 16 STR marker set in Koreans (CPI =  $8.04 \times 10^{-18}$ ) [26]. Therefore, we expect that a highly polymorphic set of 40 SNP markers with MAF > 0.45 will have nearly the same discrimination power for human identification in Koreans as would a set of 16 STR markers. In addition, 40 highly informative SNP markers with low *F<sub>st</sub>* and high heterogeneity are also more likely to prove sufficient for human identification in diverse human populations, as demonstrated in a previous study [12]. In addition, the SNP panel designed for human identification can be analyzed using different SNP genotyping methodologies [27]. In the present study, we demonstrated that 30 highly polymorphic SNP markers can distinguish up to 8,842 individuals, even in a very homogenous Korean population. Furthermore, a SNP-based human identification test identified three clinical data errors in a testing sample set (*n* = 960), including two sex information errors and one sample duplication error (data not shown). These data indicate that our SNP-based human identification method is very accurate and can be used for the quality control of biological samples in biobanks.

In this study, we have demonstrated the accuracy of SNP-based sex determination and human identification, and we have shown that SNP-based human identification test can discriminate as efficient as STR-based test when more SNP markers were used. In particular, we provided 30 highly polymorphic SNP markers as well as six sex-determining SNP markers which can be used in SNP-based human identification test. We also demonstrated that discrimination power increased 10-fold for the addition of every five SNP markers in human identification. We predict that SNP-based genetic tests will increase in popularity in the near future with the addition of SNP markers associated with various phenotypic traits. These tests will prove particularly useful in the fields of forensic science and paternity testing.

**Acknowledgments** This study was supported by a grant from the Korea Centers for Disease Control (2008-S6-E-001) and a grant from

the Ministry of Health and Welfare, Republic of Korea (01-PJ10-PG6-01GN15-001).

## References

1. Urquhart A, Kimpton CP, Downes TJ, Gill P (1994) Variation in short tandem repeat sequences—a survey of twelve microsatellite loci for use as forensic identification markers. *Int J Legal Med* 107:13–20
2. Phillips C, Salas A, Sanchez JJ, Fondevila M, Gomez-Tato A, Alvarez-Dios J, Calaza M, Casares de Cal M, Ballard D, Lareu MV, Carracedo A, The SNPforID Consortium (2007) Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. *Forensic Sci Int Genet* 1:273–280
3. Børsting C, Sanchez JJ, Hansen HE, Hansen AJ, Bruun HQ, Morling N (2008) Performance of the SNPforID 52 SNP-plex assay in paternity testing. *Forensic Sci Int Genet* 2:292–300
4. Dupuy BM, Stenersen M, Egeland T, Olaisen B (2004) Y-chromosomal microsatellite mutation rates: differences in mutation rate between and within loci. *Hum Mutat* 23:117–124
5. Huang QY, Xu FH, Shen H, Deng HY, Liu YJ, Liu YZ et al (2002) Mutation patterns at dinucleotide microsatellite loci in humans. *Am J Hum Genet* 70:625–634
6. Reich DE, Schaffner SF, Daly MJ, McVean G, Mullikin JC, Higgins JM et al (2002) Human genome sequence variation and the influence of gene history, mutation and recombination. *Nat Genet* 32:135–142
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. Kohnemann 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
9. Inagaki S, Yamamoto Y, Doi Y, Takata T, Ishikawa T, Imabayashi K et al (2004) A new 39-plex analysis method for SNPs including 15 blood group loci. *Forensic Sci Int* 144:45–57
10. Lee HY, Park MJ, Yoo JE, Chung U, Han GR, Shin KJ (2005) Selection of twenty-four highly informative SNP markers for human identification and paternity analysis in Koreans. *Forensic Sci Int* 148:107–112
11. Li L, Li CT, Li RY, Liu Y, Lin Y, Que TZ et al (2006) SNP genotyping by multiplex amplification and microarrays assay for forensic application. *Forensic Sci Int* 162:74–79
12. Pakstis AJ, Speed WC, Kidd JR, Kidd KK (2007) Candidate SNPs for a universal individual identification panel. *Hum Genet* 121:305–317
13. Sanchez JJ, Phillips C, Børsting C, Balogh K, Bogus M, Fondevila M et al (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis* 27:1713–1724
14. 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
15. Fan JB, Gunderson KL, Bibikova M, Yeakley JM, Chen J, Garcia EW, Lebruska LL, Larent M, Shen R, Barker D (2006) Illumina universal bead arrays. *Methods Enzymol* 410:57–73
16. Kidd KK, Pakstis AJ, Speed WC, Grigorenko EL, Kajuna SL, Karoma NJ et al (2006) Developing a SNP panel for forensic identification of individuals. *Forensic Sci Int* 164:20–32
17. Cho YS, Go MJ, Kim YJ, Heo JY, Oh JH, Ban HJ et al (2009) A large-scale genome-wide association study of Asian populations

- uncovers genetic factors influencing eight quantitative traits. *Nat Genet* 41:527–534
18. Rozen S, Skaletsky H (1999) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 132:365–386
  19. Nickerson DA, Tobe VO, Taylor SL (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res* 25:2745–2751
  20. Durkin AS, Cedrone E, Sykes G, Boles D, Reid YA (2000) Utility of gender determination in cell line identity. *In Vitro Cell Dev Biol Anim* 36:344–347
  21. Demichelis F, Greulich H, Macoska JA, Beroukhim R, Sellers WR, Garraway L et al (2008) SNP panel identification assay (SPIA): a genetic-based assay for the identification of cell lines. *Nucleic Acids Res* 36:2446–2456
  22. Tschentscher F, Frey UH, Bajanowski T (2008) Amelogenin sex determination by pyrosequencing of short PCR products. *Int J Legal Med* 122:333–335
  23. Shadrach B, Commane M, Hren C, Warshawsky I (2004) A rare mutation in the primer binding region of the amelogenin gene can interfere with gender identification. *J Mol Diagnostics* 6:401–405
  24. Chong SS, Kristjansson K, Cota J, Handyside AH, Hughes MR (1993) Preimplantation prevention of X-linked disease: reliable and rapid sex determination of single human cells by restriction analysis of simultaneously amplified ZFX and ZFY sequences. *Hum Mol Genet* 2:1187–1191
  25. Bouakaze C, Keyser C, Crubezy E, Montagnon D, Ludes B (2009) Pigment phenotype and biogeographical ancestry from ancient skeletal remains: inferences from multiplexed autosomal SNP analysis. *Int J Legal Med* 123:315–325
  26. Hong SB, Kim KS, Hwang JH, Kim SH, Lim SK, Lee DS et al (2006) Genetic population data on 15 autosomal STR (PowerPlex 16) loci for forensic DNA profiling in Korean. *Kor J Forensic Sci* 7:1–4
  27. Sobrino B, Brion M, Carracedo A (2005) SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Sci Int* 154:181–194