



Evaluation of ten SNP Markers for Human Identification and Paternity Analysis in Persian Population

Sajad Habibi ¹, Amirhossein Ahmadi ², Mehrdad Behmanesh ³, Ali Miri ¹, Mahmood Tavallaie ^{1,*}

¹Human Genetic Research Center, Baqiatallah University of Medical Sciences, Tehran, Iran

²Department of Genetics, Faculty of Sciences, Persian Gulf University, Bushehr, Iran

³Department of Genetics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran

* **Corresponding author:** Mahmood Tavallaie, Professor, Human Genetic Research Center, Baqiatallah University of Medical Sciences, Tehran, Iran. Fax/Tel: +98-2186034937; E-mail: mahtavalla@bmsu.ac.ir

Abstract

Background: DNA markers are inevitable tools of human identification in forensic science. Single Nucleotide Polymorphisms (SNPs) are one category of these markers which is concerned to use especially in the case of degraded DNA because of their short amplicons.

Objectives: Detection of highly informative SNPs by the criteria is the essential step to develop a useful panel of SNP markers. The purpose of this work is to get high informative SNPs for human identification in Persian ethnic of the Iranian population.

Material and Methods: Genotype and allele frequencies of 10 SNPs from the SNPforID browser were determined by a PCR-RFLP method on 100 samples that was taken from 100 unrelated Persian people.

Results: These ten SNPs were in Hardy-Weinberg equilibrium (P value > 0.1) except rs1355366 (P value = 0.02) and Heterozygosity of seven SNPs is greater than 0.45 but minor allele frequency of only four SNPs is more than 0.45. According to criteria only three SNPs rs1454361, rs2111980 and rs2107612 can pass all standards and are highly informative in population for forensic uses.

Conclusions: Our data showed that the CPI (Combined probability of Identity) and CPE (Combined Power of Exclusion) for ten SNPs are 1.13×10^{-4} and 0.809 respectively. It was also showed based on the criteria only three SNPs (rs2107612, rs1454361 and rs2111980) are highly informative in Persian population. If we can find 39 SNPs with PE and PI close to PE and PI of these three SNPs (rs2107612, rs1454361 and rs2111980), we will be able to use of these 39 SNPs in human identification with sufficient power of discrimination.

Keywords: Forensic Sciences; Forensic Anthropology; Polymorphism, Single Nucleotide

1. Background

Currently, DNA markers are common tools in paternity tests and forensic genetics (1). Short tandem repeats (STRs) as a class of well-known markers have been widely applied in forensic laboratories due to their high variability and high power of discrimination (2). However, degraded samples which recovered from disaster victims or specimens from fires, explosions and airplane crashes make STRs unsuitable for human identification due to their long amplicons (3).

Single nucleotide polymorphisms (SNPs) as the most frequent DNA sequence variations can be genotyped via short amplicons (4). However, bi-allelic SNPs are less informative than STRs, and then, more SNPs needs to achieve the same level of discrimination afforded with the STR loci used in forensic science

laboratories. Evidence showed that to obtain the same power of discrimination as 13 STRs, a panel of at least 50-100 autosomal SNP loci would be required (5). Using of highly informative SNPs with maximum power of discrimination provide a possible way to decrease the number of SNPs and overcome these limitations (6). Until now, several SNP panels have been reported to be useful for forensic tests. SNPforID browser is a database which introduces 52 SNPs with low fixation index (F_{st}) in different population for forensic application (7). Since SNP informativeness may vary significantly between populations, it is necessary to determine allele frequencies and forensic characteristics of these SNPs in each population (8).

2. Objectives

The aim of this study is to find characteristics of 10 SNPs of SNPforID database for paternity tests and find out which of them are highly informative for forensic purposes in Persian population.

3. Materials and Methods

3.1. Sample Collection

100 unrelated Persian blood samples comprised of 50 males and 50 females were collected. Written informed consent was obtained from all subjects in accordance with the declaration of Helsinki.

3.2. Genomic DNA Extraction

5 mL of peripheral blood were collected from each volunteer in test tubes containing 0.5M EDTA and DNA was extracted using DNPTM Kit (Cinnagen, Iran). Briefly, lysis solution was used to lyse blood cells and then genomic DNA from white cells selectively precipitated with isopropanol. The precipitated DNA was washed and desalted with ethanol and dissolved in TE buffer and stored in -20°C . The quantity and quality of extracted DNA were examined spectrophotometrically or visually by electrophoresis on 1% agarose gel.

3.3. Selection of Highly Informative SNP Candidates
SNP data in the SNPforID browser <http://spsmart.cesga.es/snpforid.php> were employed in the present study. In order to select 10 highly informative SNP candidates, the allelic frequencies were used for screening. As markers with even allelic distributions have high observed heterozygosity and are more informative, three of the SNPs were selected with a common 45:55– 50:50 allelic distributions in the Middle East population where Iran is located there. All of the selected markers are located on autosomal chromosomes (**Table 1**).

3.4. SNP Genotyping

Ten SNPs was determined by Polymerase Chain Reaction-restriction Fragment Length Polymorphism (PCR-RFLP). Briefly, the fragment which contains SNP were amplified to form PCR products as below: initial denaturation at 95°C for 5min, 35 cycles of 30s at 95°C , 30 s at 58 or 60°C , 30 Sec at 72°C and a final extension step of 5min at 72°C . PCR products were digested with appropriate enzymes at 37°C for overnight, according to manufacturer's instruction and then analyzed on 2% agarose gel (**Table 2**).

3.5. Statistical Analyses

The gene count method was used to calculate observed allele frequencies and the observed genotypes at each of the 10 SNP loci. probability of Identity (PI), Power of Exclusion (PE), Power of Discrimination (PD) and were calculated by PowerStatsV12.xls software, which is freely available on the net (9).

Fisher's exact tests using the OEGE program were performed to evaluate compliance with the Hardy-Weinberg equilibrium (10). The following SNP selection criteria were used to choose highly informative SNPs: [1] (MAF) > 0.45 in all ethnic groups, [2] heterozygosity > 0.45, [3] Hardy-Weinberg equilibrium (HWE) P value > 0.1, and [4] physical distance between SNP markers > 50 Mb (11).

4. Results

4.1. Genotype and Allele Frequencies of SNPs

Genotype and allele frequencies of SNP candidates in Persian population were calculated based on the gene count method (**Table 3**). All SNPs except for rs1355366 are in the Hardy-Weinberg Equilibrium (HWE) with the p-value greater than 0.1 so one marker (rs1355366) is not consistent with HWE in population under study. Heterozygosity of seven SNPs is greater than 0.45 but minor allele frequency of only four SNPs is more than 0.45 (Of course, rs1454361 is very close to (MAF) > 0.45). Altogether according to four criteria only three SNPs of rs1454361, rs2111980 and rs2107612 can pass all standards and are highly informative in population for forensic uses.

4.2. Calculation of Combined PI and PE of Ten SNPs in Persian Population

PI and PE were calculated for each of candidate SNPs. With 10 autosomal SNPs combined PI (CPI) and combined PE (CPE) in Persian population was calculated as 1.13×10^{-4} and 0.80900 respectively (**Table 3**).

5. Discussion

Since geneticists discovered polymorphic structures in human genomes, several genetic tests have been developed to utilize this feature in paternity tests and very soon became profitable tools for forensic applications (12). Although STRs are widely employed in genetic tests as suitable tools; SNPs have some advantages in comparison to STRs. Short amplicon size, low mutation rates and genotyping via high-throughput technologies are some of their advantages (13, 14).

A panel of 52 SNPs which are presented in SNPforID browser is suggested to be enough informative and has the same power of discrimination as 15 STRs (15, 16). However, the SNP informativeness may vary significantly between populations, so, genotype and allele frequencies of 10 SNPs of this database were calculated in Persian population.

As shown in (**Table 3**) some of these SNPs have different distributions in the Persian population in comparison with those reported for Middle East population in SNPforID browser (7).

Table 1. Primers and characteristics of amplified human genomic DNA segments containing selected SNP markers

NCBI SNP cluster ID	PCR product (bp)	Annealing temperature	Primers		Chromosomal location	Chromosome
rs2107612	370	60 °C	5'-TCAGGGAGGAATAAACATTACAGG-3'	5'-GTCACAGCAACATAACCATATTAC-3'	888,320	12
rs1454361	412	60 °C	5'-ACTCTAGTGACATAGCCTCCAGTG-3'	5'-TTGGAAGGACAATGAAGTTGCACG-3'	25,850,832	14
rs2111980	305	60 °C	5'-CTCTTAACCTCCCTCCCTTGCGTG-3'	5'-CTCTTCCCTTCGGCCCACTCCAAC-3'	106,328,254	12
rs1355366	400	60 °C	5'-CCGTCTCTTCAGGAACAGTAGAAC-3'	5'-CAGGTGTGACAGCCCTAGTTCTG-3'	190,806,108	3
rs251934	230	60 °C	5'-ACTGACCCTTGACAGAACTGACC-3'	5'-CCCAAGTCCATCATTGGCTGT-3'	174,778,678	5
rs1028528	614	58 °C	5'-TCCCTCAGCAAGAATCCCATTAGG-3'	5'-TCAGTACAACCCTGCAAGATAGATG-3'	48,362,290	22
rs1335873	476	60 °C	5'-GGAATGGGTCAGGTCGAAGTCC-3'	5'-AAGAACAGGAGCGTCAGAAACAG-3'	20,901,724	13
rs1979255	666	60 °C	5'-AAGAGAAGATACAGAGGCAATTCAG-3'	5'-CTCCCTTGACATCCCAAAGCATACC-3'	190,318,080	4
rs917118	711	60 °C	5'-CTGATTGTCTAGTGGCAGCGTTC-3'	5'-GTGTCCAGCAAGAGAGATTTC-3'	4,457,003	7
rs1413212	655	58 °C	5'-AGTGTAAAGTGAATTTGCCCTATGCC-3'	5'-CACAAACCTAAGACTTGCCTTCAG-3'	24,806,797	1

Table 2. Restriction enzymes which were used for genotyping by RFLP-PCR method

SNP	Enzyme	Restriction site	Cut Allele	Fragment size after digestion (bp)
rs2107612 (A/G)	XmiI	GTMKAC	A	205 (165)
rs1454361 (A/T)	XceI	RCATGY	A	212 (200)
rs2111980 (A/G)	SacI	GAGCTC	G	200 (105)
rs1355366 (A/G)	FaqI	GGGAC(10/14N)	G	216(184)
rs251934 (C/T)	XmiI	GTMKAC	C	144(90)
rs1028528 (A/G)	DraIII	CACNNNGTG	A	340(274)
rs1335873 (A/T)	XmiI	GTMKAC	T	256(220)
rs1979255 (C/G)	AlwI	GGATC	C	406(260)
rs917118 (C/T)	HgaI	GACGC	C	431(280)
rs1413212 (T/C)	BsmFI	GTCCC	T	437(218)

Table 3. Paternity index, probability of exclusion, genotype and allele frequencies of ten SNPs in Persian population

NCBI SNP cluster ID	PI	PE	PD	Allele Frequency						Genotype Frequency			HWE P-value
				Middle East		Persian				Homo allele 1	Het allele 1,2	Homo allele 2	
				Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2				
rs2107612	0.391	0.215	0.609	G	A	0.33	0.67	0.495	0.505	0.23	0.53	0.24	0.54
rs1454361	0.399	0.216	0.601	T	A	0.54	0.46	0.443	0.557	0.18	0.53	0.29	0.46
rs2111980	0.362	0.158	0.638	A	G	0.48	0.51	0.535	0.465	0.3	0.46	0.24	0.44
rs1355366	0.342	0.102	0.658	A	G	0.62	0.38	0.55	0.45	0.36	0.38	0.26	0.02
rs251934	0.439	0.140	0.561	T	C	0.58	0.42	29	71	0.07	0.44	0.49	0.49
rs1028528	0.465	0.113	0.535	G	A	0.64	0.36	25	75	0.05	40	0.55	0.5
rs1335873	0.422	0.147	0.578	A	T	0.38	0.62	0.315	0.685	0.09	0.45	0.46	0.66
rs1979255	0.43	0.154	0.57	G	C	0.29	0.71	0.31	0.69	0.08	0.46	0.46	0.45
rs917118	0.365	0.14	0.635	T	C	0.59	0.41	0.425	0.575	0.2	0.45	0.35	0.42
rs1413212	0.433	0.133	0.567	G	A	0.31	0.69	0.645	0.355	0.38	0.53	0.09	0.11
Total PI								1.13 E-04					
Total PE								0.809000					

These differences could alter forensic parameters of SNPs for forensic application (8). Based on the allele frequencies of 10 SNPs, CPI was calculated as 1.13E-04. This is very more than 2.2E-17 which are supplied with 15 STR loci in the Persian population, so we should use of vey more SNPs for equal CPI with 15 STR but if selected SNPs are high informative by criteria we can use of fewer number of SNPs. This is same about CPE that for equal power of exclusion with 15 STRs we should use very more SNP markers (17). These data mentioned that addition of more appropriate SNPs must be necessary to increase the power of discrimination. Gill reported that 50–80 loci are taken to reach the same discrimination level of 16 STRs loci (18). The addition of more SNPs, however, is associated with some restrictions. Production and commercialization of robust assays containing large numbers of oligonucleotide PCR primers will not be

trivial (3).

Furthermore, examining more loci will be more expensive and time-consuming; therefore, introducing highly informative SNPs is the unavoidable direction in order to decrease the number of SNPs for forensic uses (11). Since most SNPs are biallelic, the maximum power of discrimination is taken by a combination of SNPs with a nearly the same distribution between two alleles (19).

Based on this procedure, Lee et al introduced 24 highly informative SNPs in Korean population with a PI and PE of 1.9E-10 and 0.989 respectively which is corresponding to nine STR loci (8).

Kim et al suggested that a highly polymorphic set of about 40 SNP markers with MAF > 0.45 and the even allelic frequency will have nearly the same discrimination power for human identification as would a set of 16 STR markers in Korean population (11).

6. Conclusions

Here in Persian population, it is shown that three out of ten SNPs (rs1454361, rs2111980 and rs2107612) are highly informative based on criteria. With the average PI (0.387) and PE (0.197) of these three SNPs, it is estimated that 39 SNPs have CPI and CPE of 8.33E-17 and 0.9998 respectively, which is comparable to 15 STRs in Persian population with a CPI of 2.2E-17 and a CPE of 0.9999 (17). These numbers of SNPs is less than that 50-80 reported to be necessary for human identification.

Altogether, although it is a preliminary study with a small number of SNPs, the outcomes of this study help forensic labs to define a list of common highly informative SNPs which are applicable in all populations.

Acknowledgments

The authors thank the Research Section of Human Genetic Research Center of Baqiatallah University and Department of Molecular Genetics of Tarbiat Modares University for their help and guidance on this study.

References

- Roper SM, Tatum OL. Forensic aspects of DNA-based human identity testing. *J Forensic Nurs.* 2008;**4**(4):150-156. [pmid: 19418771](#)
- Lygo JE, Johnson PE, Holdaway DJ, Woodroffe S, Whitaker JP, Clayton TM, et al. The validation of short tandem repeat (STR) loci for use in forensic casework. *Int J Legal Med.* 1994;**107**(2):77-89. [pmid: 7819123](#)
- Butler JM, Coble MD, Vallone PM. STRs vs. SNPs: thoughts on the future of forensic DNA testing. *Forensic Sci Med Pathol.* 2007;**3**(3):200-205. [doi: 10.1007/s12024-007-0018-1](#) [pmid: 25869164](#)
- Sobrino B, Brion M, Carracedo A. SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Sci Int.* 2005;**154**(2-3):181-194. [doi: 10.1016/j.forsciint.2004.10.020](#) [pmid: 16182964](#)
- Budowle B, van Daal A. Forensically relevant SNP classes. *Biotechniques.* 2008;**44**(5):603-608, 610. [doi: 10.2144/000112806](#) [pmid: 18474034](#)
- Boonyarit H, Mahasirimongkol S, Chavalvechakul N, Aoki M, Amitani H, Hosono N, et al. Development of a SNP set for human identification: A set with high powers of discrimination which yields high genetic information from naturally degraded DNA samples in the Thai population. *Forensic Sci Int Genet.* 2014;**11**:166-173. [doi: 10.1016/j.fsigen.2014.03.010](#) [pmid: 24747184](#)
- Amigo J, Phillips C, Lareu M, Carracedo A. The SNPforID browser: an online tool for query and display of frequency data from the SNPforID project. *Int J Legal Med.* 2008;**122**(5):435-440. [doi: 10.1007/s00414-008-0233-7](#) [pmid: 18491122](#)
- Lee HY, Park MJ, Yoo JE, Chung U, Han GR, Shin KJ. Selection of twenty-four highly informative SNP markers for human identification and paternity analysis in Koreans. *Forensic Sci Int.* 2005;**148**(2-3):107-112. [doi: 10.1016/j.forsciint.2004.04.073](#) [pmid: 15639604](#)
- Martin P, Garcia O, Heinrichs B, Yurrebaso I, Aguirre A, Alonso A. Population genetic data of 30 autosomal indels in Central Spain and the Basque Country populations. *Forensic Sci Int Genet.* 2013;**7**(2):e27-30. [doi: 10.1016/j.fsigen.2012.10.003](#) [pmid: 23138099](#)
- Rodriguez S, Gaunt TR, Day IN. Hardy-Weinberg equilibrium testing of biological ascertainment for Mendelian randomization studies. *Am J Epidemiol.* 2009;**169**(4):505-514. [doi: 10.1093/aje/kwn359](#) [pmid: 19126586](#)
- Kim JJ, Han BG, Lee HI, Yoo HW, Lee JK. Development of SNP-based human identification system. *Int J Legal Med.* 2010;**124**(2):125-131. [doi: 10.1007/s00414-009-0389-9](#) [pmid: 19921517](#)
- Luftig MA, Richey S. DNA and forensic science. *New Eng Rev.* 2000;**35**:609.
- Kidd KK, Pakstis AJ, Speed WC, Grigorenko EL, Kajuna SL, Karoma NJ, et al. Developing a SNP panel for forensic identification of individuals. *Forensic Sci Int.* 2006;**164**(1):20-32. [doi: 10.1016/j.forsciint.2005.11.017](#) [pmid: 16360294](#)
- Phillips C, Fondevila M, Garcia-Magarinos M, Rodriguez A, Salas A, Carracedo A, et al. Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. *Forensic Sci Int Genet.* 2008;**2**(3):198-204. [doi: 10.1016/j.fsigen.2008.02.002](#) [pmid: 19083821](#)
- Moreno F, Freire-Aradas A, Phillips C, Fondevila M, Carracedo A, Lareu MV. SNP variation with latitude: Analysis of the SNPforID 52-plex markers in north, mid-region and south Chilean populations. *Forensic Sci Int Genet.* 2014;**10**:12-16. [doi: 10.1016/j.fsigen.2013.12.009](#) [pmid: 24680124](#)
- Musgrave-Brown E, Ballard D, Balogh K, Bender K, Berger B, Bogus M, et al. Forensic validation of the SNPforID 52-plex assay. *Forensic Sci Int Genet.* 2007;**1**(2):186-190. [doi: 10.1016/j.fsigen.2007.01.004](#) [pmid: 19083753](#)
- Shepard EM, Herrera RJ. Iranian STR variation at the fringes of biogeographical demarcation. *Forensic Sci Int.* 2006;**158**(2-3):140-148. [doi: 10.1016/j.forsciint.2005.05.012](#) [pmid: 15998573](#)
- Gill P. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *Int J Legal Med.* 2001;**114**(4-5):204-210. [pmid: 11355396](#)
- Krawczak M. Informativity assessment for biallelic single nucleotide polymorphisms. *Electrophoresis.* 1999;**20**(8):1676-1681. [doi: 10.1002/\(SICI\)1522-2683\(19990101\)20:8](#) [pmid: 10435431](#)