Evaluation of Informative SNPs in Iranian Azeri Population

Ali Mohammadi¹, Maryam Ramezani¹, Sayed Mostafa Hosseini¹, Mahmood Tavallai²

¹Human Genetics Research Center, Baqiyatallah University of Medical Sciences Tehran, Iran

Corresponding Author: Mahmood Tavallai, Human Genetics Research Center, Baqiyatallah University of Medical Sciences Tehran, Iran. Tell: +98-2186034956, Email: tavalla.mah@gmail.com

Received March 26, 2018; Accepted July 3, 2018; Online Published September 30, 2018

Abstract

Introduction: Two of the most important tests used in the forensic genetics are DNA fingerprinting and paternity testing. Short tandem repeats (STRs) are very frequently used in identification. Nevertheless, there are limitations on the use of these markers in identifying badly degraded DNA. In these cases, single nucleotide polymorphisms (SNPs) are used because of their shorter amplicon lengths. Recognition of SNPs of high informativeness is a vital step in preparing a list of suitable SNPs. The present research aimed to determine the maximum informative SNPs to be used for identification in Iranian Azeri population.

Materials and Methods: Four SNPs developed by the SNPforID Consortium were selected. The allele frequencies of the SNPs were obtained using HRM Analysis on DNA samples taken from 100 different individuals. The SNPs that satisfied the Hardy-Weinberg equilibrium and had heterozygosity of higher than 0.50 and equal allele frequency were selected as SNPs of the maximum informativeness.

Results: Two polymorphisms (Rs2107612 and Rs1355366) had equal allele frequencies and heterozygosity of higher than 50 percent. Therefore, these 2 polymorphisms are considered highly informative among the studied Azeri population and can be considered in preparing a list of suitable SNPs.

Conclusions: Results of the present study can be used along with other SNPs to increase the identification power for some samples. It can also help preparing a database of suitable SNPs to be used for identification in Azeri population in Iran.

Keywords: Identification, STRs, SNPs, HRM Analysis, Allele Frequencies, Heterozygosity


Introduction

New DNA analysis techniques have made advancements in human identification and forensic genetics. Using identification methods based on DNA analysis is necessary for identifying human-remains in mass disasters, wars, and sociopolitical events when personal factors or other physical traits are not informative.¹ Using short tandem repeat (STR) markers is now a common method for identification in forensic tests. Nevertheless, they cannot be used for DNA that has been considerably degraded because the standard STRs have long repetitive regions and do not yield qualified results for degraded samples. Then, alternative methods must be used in such cases.¹-⁴ One of them is genotyping mitochondrial DNA using short genetic markers. Although this method has advantages, the sequence diversity of these markers is lower in comparison with the STRs. Also, since inheritance of these markers is maternal, finding some relationships such as those between fathers and daughters is difficult which limits their capability.²⁻³ Another alternative method is to shorten fragment length by redesigning currently known STR primers that are generally called mini-STRs.³⁻⁴ Next common alternative marker in forensic tests is single nucleotide polymorphisms (SNPs) that has advantages over other markers.⁴⁻⁵ However, the prerequisite to using SNP genotyping for forensic purposes is to obtain information related to allelic and genotypic frequencies in Iranian populations to serve as the base for expanding systems possessing maximum power of discrimination.⁵ The present research was conducted to satisfy the same need. SNPs genotyping is performed via several methods. High-resolution melting (HRM) is one of the RT polymerase chain reaction (PCR) -based methods that uses differences in melt curves to recognize variations in nucleic acid sequences. The melt curve can be recognized by fluorescent dyes bound to the nucleic acids based on the separation of these dyes during denaturation of double-strand DNA by real-time PCR machine.⁶⁻⁷ Various features such as GC content, segment length, sequence, and heterozygosity cause differences in melt curves, and the results of this data can be used in SNPs genotyping, mutation screening, methylation, and other applied research.⁸⁻¹² In the SNP genotyping method, nucleotide variations cause differences in melt curves, although these differences are slight in point mutations and SNPs, as these temperature differences are from 0.5- 1°C for the G/T, G/A, C/T and C/A, 0.5-1°C for G/C, and less than 0.2°C for T/A nucleotide differences. In HRM, this temperature difference
between nucleotides appears in separate diagrams with
different colors that can be used in mutation screening and
SNP genotyping. Identification informative SNPs for any
population is a vital step in forensic genetics.

The present research intended to study allele frequency and
efficiency of the 4 SNPs introduced in the SNPforID database
(i.e. Rs1454361, Rs1355366, Rs2107612, Rs2111980).

Materials and Methods
In order to prevent the possible contamination with modern
dNA, laboratory equipment's were decontaminated through
bleaching with 5% active chlorine. Then, they were washed
in sterile distilled water, and finally, all surfaces were exposed
to UV irradiation in an ultraviolet chamber for 60 minutes.
All buffers and materials were autoclaved and sterilized. All
steps were taken in separate places under sterile conditions
using sterile gloves and masks. Negative and positive controls
were used for all stages of extraction, PCR, real-time PCR,
and sequencing.

According to the approximate distribution of ethnic groups
in Iran's population, one hundred unrelated individuals that
lived in Northwest regions of Iran (40 from East Azerbaijan
province, 25 from West Azerbaijan province, 20 from Ardabil
province, and 15 from Zanjan province) were selected and
4 mL blood sample was taken from each one. The blood
samples were put in tubes containing 0.5M anticoagulant
EDTA at pH=8 and transferred to the laboratory.

Rapid genomic DNA extraction (RGDE) method was used
to extract DNA from blood samples. By this method, genomic
DNA with high quality and quantity can be obtained in the
shortest time. Then electrophoresis by 1% agarose gel was
performed for quality control of the DNA extraction product.
Moreover, DNA quantity was controlled using NanoDrop
(IMPLEN Company), and samples with low quality and
quantity were re-extracted.

The nested PCR method was employed to obtain better and
more accurate results. In this method, 2 pairs of primers are
used to increase PCR sensitivity (Tables 1 and 2). All Primers
were designed using Oligo7 software. The first pair is used
to amplify specific segments of the target DNA for 30 cycles
(Table 3). Then PCR product was transferred to another
tube and used as template, then HRM was performed by the
second primer pairs (Table 4) (StepOnePlus™ Real-Time
PCR System by Thermo Fisher). Then, the serial dilution of
the initial PCR product was prepared, and HRM analysis was
carried out for each one. It was found that the 0.001 dilution
was the best. In the melt curve stage, the temperature was
increased from 60 degrees to 95 degrees at intervals of 0.3 C,
which was suitable for detection of SNPs.

Finally, allele frequencies of the SNPs were obtained by
HRM analysis on DNA samples taken from 100 different
individuals. At the end of the HRM analysis, sequencing was
performed by genetic analyzer machine (Applied Biosystems
3130XL) for 3 samples of each peak to confirm the final
results. The sequencing results were analyzed using the DNA
Baser and Gene Runner software and generalized to the other
samples. After genotyping samples from 100 individuals,
results for each of the 4 SNPs were statistically studied using
the PowerStats version 12 software.

Ethical Considerations
All procedures performed in studies involving human
participants accorded with the ethical standards of the
institutional and/or national research committee and with
the 1964 Helsinki Declaration and its later amendments or
comparable ethical standards (Ethics Code for Research:

Table 1. Primers Used in the Primary PCR and the Characteristics of Proliferated Fragments Containing SNP Marker

<table>
<thead>
<tr>
<th>NCBI SNP Cluster ID</th>
<th>Product Length</th>
<th>Annealing Temperature</th>
<th>Primers Used in Primary PCR</th>
<th>SNP Genomic Position</th>
<th>Chromosome</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2107612</td>
<td>370</td>
<td>60°C</td>
<td>5'-TACAGGCTGACACATATTACGAG-3'</td>
<td>888320</td>
<td>12</td>
</tr>
<tr>
<td>rs1454361</td>
<td>412</td>
<td>60°C</td>
<td>5'-ACGTGTGACAGCCTAGTTCTG-3'</td>
<td>25850832</td>
<td>14</td>
</tr>
<tr>
<td>rs2111980</td>
<td>305</td>
<td>60°C</td>
<td>5'-TCCTACTTCCCCCTTGCGCTG-3'</td>
<td>106328254</td>
<td>12</td>
</tr>
<tr>
<td>rs1355366</td>
<td>400</td>
<td>60°C</td>
<td>5'-CCGCTCTCTACAGGAACAGTACA-3'</td>
<td>190806108</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 2. Primers Used in HRM and the Characteristics of Proliferated Fragments Containing SNP Marker

<table>
<thead>
<tr>
<th>NCBI SNP Cluster ID</th>
<th>Product Length</th>
<th>Annealing Temperature</th>
<th>Primers used in Primary PCR</th>
<th>SNP Genomic Position</th>
<th>Chromosome</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2107612</td>
<td>109</td>
<td>57°C</td>
<td>5-GACATTCATCTCCTGGTTAA-3</td>
<td>888320</td>
<td>12</td>
</tr>
<tr>
<td>rs1454361</td>
<td>72</td>
<td>57°C</td>
<td>5-TGGGAGGGAAGGAAATAC-3</td>
<td>25850832</td>
<td>14</td>
</tr>
<tr>
<td>rs2111980</td>
<td>77</td>
<td>57°C</td>
<td>5-ATTACGCTGATTCAGGAACAGTAC-3</td>
<td>106328254</td>
<td>12</td>
</tr>
<tr>
<td>rs1355366</td>
<td>64</td>
<td>57°C</td>
<td>5-CAATTTCAGACCCACTGGGAG-3</td>
<td>190806108</td>
<td>3</td>
</tr>
</tbody>
</table>
Results

After DNA extraction, results were assayed with electrophoresis (Figure 1). Electrophoresis showed that DNA extraction has been successful. Following the HRM analysis for each SNP, diagrams indicated segregation and identification of the SNPs for each individual. Each SNP is represented by a specific color (Figures 2-4).

Difference plot showed that the HRM has been done successfully. Allele frequency for Rs2107612 was G: 0.389, A: 0.611; for Rs1454361 was A: 0.789, T: 0.211; for Rs1355366 was T: 0.505, C:495; and for Rs2111980 was G: 0.921, and A: 0.079.

Then data was entered into the PowerStats version 12 software and various indicators listed in the following table (Table 5). Based on the results summarized in Table 5, minimum MP was determined for Rs2107612, and also,

<table>
<thead>
<tr>
<th>Index</th>
<th>Rs2107612</th>
<th>Rs1454361</th>
<th>Rs1355366</th>
<th>Rs2111980</th>
</tr>
</thead>
<tbody>
<tr>
<td>MP</td>
<td>0.414</td>
<td>0.579</td>
<td>0.594</td>
<td>0.734</td>
</tr>
<tr>
<td>PD</td>
<td>0.586</td>
<td>0.421</td>
<td>0.406</td>
<td>0.266</td>
</tr>
<tr>
<td>PE</td>
<td>0.212</td>
<td>0.009</td>
<td>0.510</td>
<td>0.019</td>
</tr>
<tr>
<td>TPI</td>
<td>1.06</td>
<td>0.56</td>
<td>2.00</td>
<td>0.59</td>
</tr>
<tr>
<td>Ho</td>
<td>47.4%</td>
<td>89.5%</td>
<td>25.0%</td>
<td>84.2%</td>
</tr>
<tr>
<td>He</td>
<td>52.6%</td>
<td>10.5%</td>
<td>75.0%</td>
<td>15.8%</td>
</tr>
<tr>
<td>A1 Frq</td>
<td>38.9%</td>
<td>21.1%</td>
<td>50.5%</td>
<td>92.1%</td>
</tr>
<tr>
<td>A2 Frq</td>
<td>61.1%</td>
<td>78.9%</td>
<td>49.5%</td>
<td>7.9%</td>
</tr>
</tbody>
</table>

MP (Matching Probability): This index indicates identical likely this SNP for 2 unrelated individuals in the studied population. PD (power of disclosure): This index indicates the power of disclosure for 2 unrelated individuals in the studied population by this SNP PE (power of exclusion): Probability of excluding relatives of the true father from paternity. TPI (Typical Paternity Index): Is a calculated value generated for a single genetic marker or locus and is associated with the statistical strength or weight of that locus in favor of or against parentage. Ho (homozygosity): possessing 2 identical forms of a particular gene, one inherited from each parent. He (heterozygosity): possessing 2 different forms of a particular gene, one inherited from each parent. A1 Frq (Allele 1 Frequency): Is the relative frequency of allele 1 at a particular locus in the studied population. A2 Frq (Allele 2 Frequency): Is the relative frequency of allele 2 at a particular locus in the studied population.
maximum of PD was determined for Rs2107612 which indicates this SNP has greater power of discrimination for 2 unrelated individuals in the studied population.

**Discussion**

Iran can be considered as a major route for the migration of humans from Africa to Southwest Asia as it is located in a strategic location in West Asia where 3 continents of Asia, Europe, and Africa are joined. Identification of autosomal and mtDNA SNPs in different ethnic groups in Iran can show the relationship between Iranian ethnicities and other ethnicities, and could be informative in determining the pattern of human migration. Since most SNPs have 2 alleles, their maximum discriminatory power is obtained when a set of them that have equal allelic frequency are used together. Informativeness of SNPs may vary in different populations. For the present research 4 SNPs were selected from the SNPforID database. They had 50:50 allelic frequencies in various populations so it could repeat in other population. Allele frequencies of the SNPs were obtained by using HRM analysis from 100 different individuals that lived in Northwest of Iran (East Azerbaijan province, West Azerbaijan province, Ardabil province, and Zanjan province). According to our findings, only 2 polymorphisms (Rs2107612 and Rs1355366) had equal allelic frequencies with heterozygosity more than 50%. Accordingly, these polymorphisms are very informative and can be used to identify Iranian Azeri population.

Also, the minimum MP (matching probability) was determined for Rs2107612, and the maximum PD (Power of disclusion) for Rs2107612; which indicates this SNP has greater power of discrimination for 2 unrelated individuals in the studied population.

Furthermore, comparison of allelic frequencies of all 4 SNPs in the currently studied population with other studied populations base of SNPforID database showed that allelic frequencies of these 4 SNPs were very similar to those of the Persian speaking populations in Iran (Figures 5-8). This finding is in line with the previous studies that showed the genetic distances between Iranians and populations in Central Asia, East Asia, and Southeast Asia were higher than those between Iranians and populations in West Asia such as Turkey and the Caucasus.
Conclusions
Recognition of SNPs of high informativeness is a vital step in preparing a list of suitable SNPs for identification in any population. The present research aimed to identify the most informative SNPs owing potential to be benefited for identification in Iranian Azeri population. Based on our study, SNPs Rs2107612 and Rs1355366 could be included in the list of suitable SNPs. Results of the present study can be used along with other SNPs to increase the identification power for some samples, and it can also help to preparing a database of suitable SNPs to be used for identification in Iranian Azeri population.

Authors’ Contributions
AM developed the theoretical formalism, carried out the experiments, performed the analytic calculations and wrote the manuscript; MR prepared the samples; SMH conceived the original idea; and MT supervised the project.

Conflict of Interest Disclosures
The authors declare they have no conflicts of interest.

http://www.biotechrep.ir

Acknowledgments
We would like to thank personnel of the NOOR Humans Genetics Research Center (Ebrahim Kiani, Hasan Bahmani, Sajad Habibi, Ali Miri, Peyman Zargari) for all their support of researchers.

References
6. Twist GP, Gaedigk R, Leeder JS, Gaedigk A. High-resolution melt


