J Appl Biotechnol Rep. 2018 Sep;5(3):125-130

doi 10.29252/jabr.05.03.07

Journal of
Applied Biotechnology
Reports

Original Article

Evaluation of Informative SNPs in Iranian Azeri Population

Ali Mohammadi[1], Maryam Ramezani[1], Sayed Mostafa Hosseini[1], Mahmood Tavallai[1*]

[1]Human 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 Azari 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

Citation: Mohammadi A, Ramezani M, Hosseini SM, Tavallai M. Evaluation of informative SNPs in Iranian Azeri population. J Appl Biotechnol Rep. 2018;5(3):125-130. doi:10.29252/jabr.05.03.07.

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.[6,7] 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 C/G, 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.[6-8] Identification informative SNPs for any population is a vital step in forensic genetics.[5-8]

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.[9] 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 Primitive PCR and the Characteristics of Proliferated Fragments Containing SNP Marker

NCBI SNP Cluster ID	Product Length	Annealing Temperature	Primers Used in Primary PCR	SNP Genomic Position	Chromosome
rs2107612	370	60°C	5′- TCAGGGAGGAATAAACATTACAGG-3′ 5′- GTCACAGCAACATAACCATATTAC-3′	888320	12
rs1454361	412	60°C	5′-ACTCTAGTGACATAGCCTCCAGTG-3′ 5′-TTGGAAGGACAATGAAGTTGCACG-3′	25850832	14
rs2111980	305	60°C	5′- CTCTTAACCTCCCTCCCTTGCCTG-3′ 5′- CTCTTCCTTCCGCCCCACTCCAAC-3′	106328254	12
rs1355366	400	60°C	5′- CCGTCTCTTCAGGAACAGTAGAAC-3′ 5′- CAGGTGTGACAGCCTAGTTCTG-3′	190806108	3

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

NCBI SNP Cluster ID	Product Length	Annealing Temperature	Primers used in Primary PCR	SNP Genomic Position	Chromosome
rs2107612	109	57 °C	5-GAGCATTCTCTTCTGTTAA-3 5-TACATTATTCAACTGTTTTGG-3	888320	12
rs1454361	72	57 °C	5-TGGGGAGGAGGGAAATAC-3 5-GCTGTCCATCATCAGTAAGAC-3	25850832	14
rs2111980	77	57 °C	5-TTTACGGTCAAAGCATCTT-3 5-ATCTTTGCCAGTGAGTCA-3	106328254	12
rs1355366	64	57 °C	5-AATTTCAGAGCCACTGGAG-3 5-GTGCTTAGGCCACAACTG-3	190806108	3

Table 3. Primary PCR was done using Taq DNA Polymerase 2x Master Mix (Ampliqon) with the following conditions

Denaturation	95°C	30 seconds
Annealing	60°C	30 seconds
Extension	72°C	30 seconds
Final Extension	72°C	30 seconds

Table 4. HRM was performed using StepOne Plus machine (ABI) and 5x Hot FIREPOL Evagreen HRM mix-Rox (Solis BioDyne) with the following conditions

Denaturation	95°C	20 seconds
Annealing	57°C	25 seconds
Extension	60°C	20 seconds
Melt curve	95°C	10 seconds
Melt curve	60°C	60 seconds
Melt curve	95°C	15 seconds

IR.BMSU.RBC.1396.177)

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.

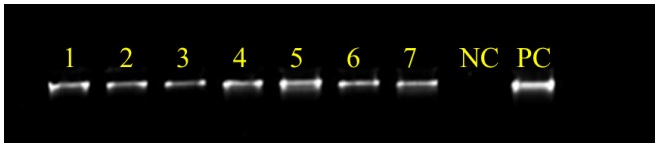

Figure 1. Agarose Gel Electrophoresis for 7 Extracted DNA With Negative and Positive Control.

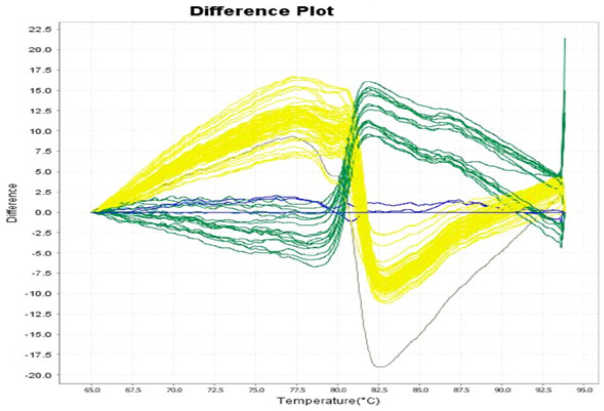

Figure 2. Difference Plot Obtained by HRM Analysis. Each peak represents an SNP.

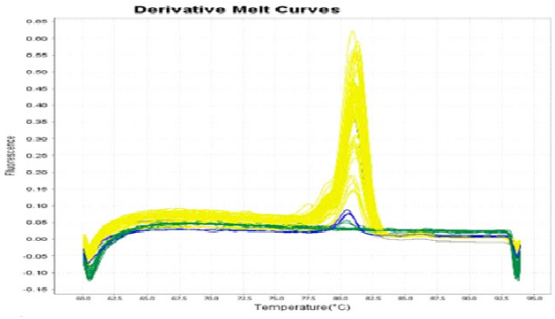

Figure 3. Derivative Melt Curves Obtained by HRM Analysis. Each peak represents an SNP.

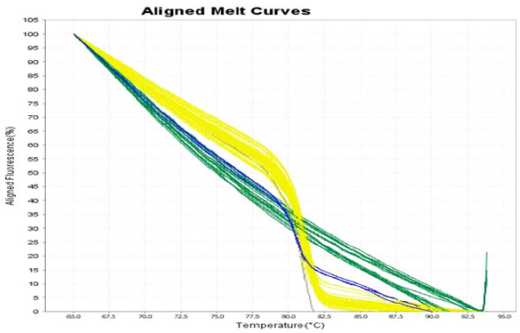

Figure 4. Aligned Melt Curves Obtained by HRM Analysis. Each peak represents an SNP.

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 5. Various Indexes Based on PowerStats

Index	Rs2107612	Rs1454361	Rs1355366	Rs2111980
MP	0.414	0.579	0.594	0.734
PD	0.586	0.421	0.406	0.266
PE	0.212	0.009	0.510	0.019
TPI	1.06	0.56	2.00	0.59
Ho	47.4%	89.5%	25.0%	84.2%
He	52.6%	10.5%	75.0%	15.8%
A 1 Frq	38.9%	21.1%	50.5%	92.1%
A 2 Frq	61.1%	78.9%	49.5%	7.9%

MP (Matching Probability): This index indicates identical likely this SNP for 2 unrelated individuals in the studied population. PD (power of disclusion): This index indicates the power of disclusion 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. A 1 Frq (Allele 1 Frequency): Is the relative frequency of allele 1 at a particular locus in the studied population. A 2 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.[10-13] 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.[13,18] 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.[16-20] Informativeness of SNPs may vary in different populations.[13-20] 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.[13,18]

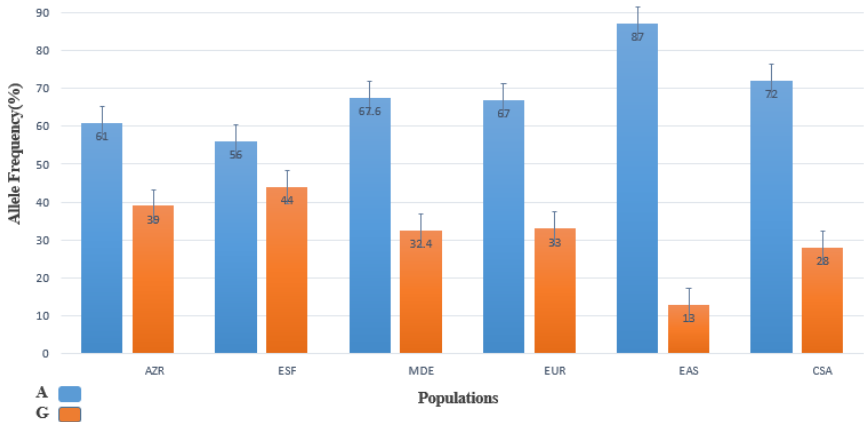

Figure 5. Allele Frequencies Comparison for Rs2107612 Loci Among the Iranian Azeri Population and others Population. Esfahan Fars (ESF); Middle East (MDE); Europe (EUR); East Asia (EAS); Southeast Asia (SEA).

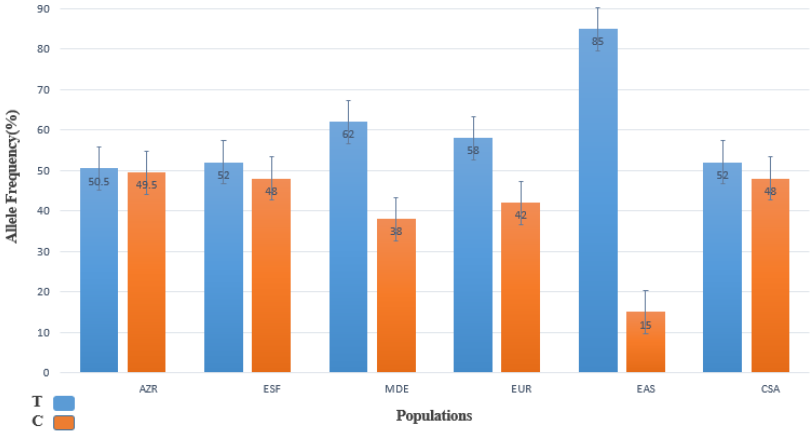

Figure 6. Allele Frequencies Comparison for Rs1355366 Loci Among the Iranian Azeri Population and Others Population. Esfahan Fars (ESF); Middle East (MDE); Europe (EUR); East Asia (EAS); Southeast Asia (SEA).

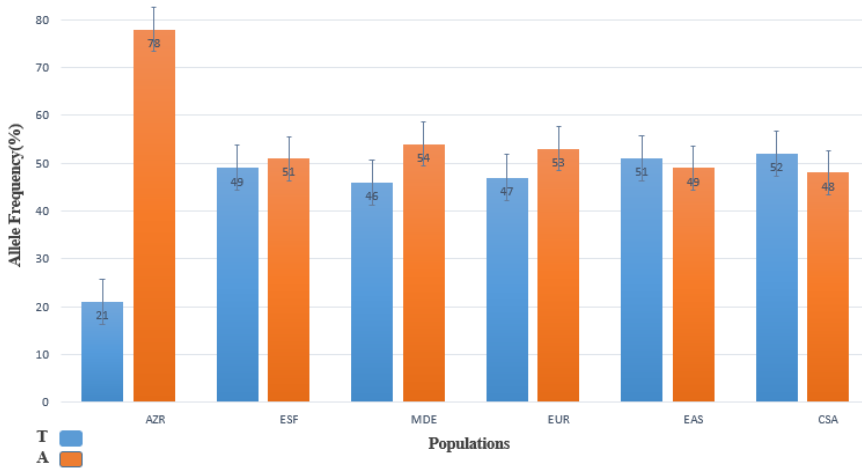

Figure 7. Allele Frequencies Comparison for Rs1454361 Loci Among the Iranian Azeri Population and Others Population. Esfahan Fars (ESF); Middle East (MDE); Europe (EUR); East Asia (EAS); Southeast Asia (SEA).

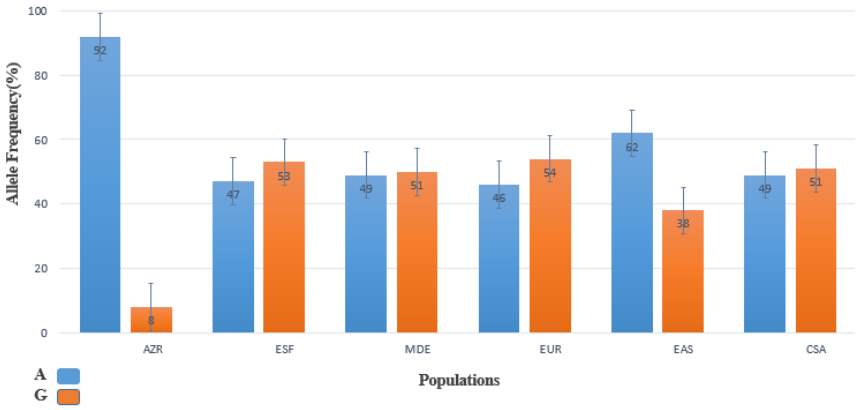

Figure 8. Allele Frequencies Comparison for Rs2111980 Loci Among the Iranian Azeri Population and Others Population. Esfahan Fars (ESF); Middle East (MDE); Europe (EUR); East Asia (EAS); Southeast Asia (SEA).

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.

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

1. Mohammadi A, Ghorbani Alvanegh A, Khafaei M, et al. A New and Efficient Method for DNA Extraction from Human Skeletal Remains Usable in DNA Typing. J Appl Biotechnol Rep. 2017;4(2):609-614.
2. 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.
3. Amorim A, Pereira L. Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. Forensic Sci Int. 2005;150(1):17-21. doi:10.1016/j.forsciint.2004.06.018.
4. Kayser M, de Knijff P. Improving human forensics through advances in genetics, genomics and molecular biology. Nat Rev Genet. 2011;12(3):179-192. doi:10.1038/nrg2952.
5. Farjadian S, Sazzini M, Tofanelli S, et al. Discordant patterns of mtDNA and ethno-linguistic variation in 14 Iranian Ethnic groups. Hum Hered. 2011;72(2):73-84. doi:10.1159/000330166.
6. Twist GP, Gaedigk R, Leeder JS, Gaedigk A. High-resolution melt

analysis to detect sequence variations in highly homologous gene regions: application to CYP2B6. Pharmacogenomics. 2013;14(8):913-922. doi:10.2217/pgs.13.66.
7. Wang J, Pan X, Liang X. Assessment for Melting Temperature Measurement of Nucleic Acid by HRM. J Anal Methods Chem. 2016;2016:5318935. doi:10.1155/2016/5318935.
8. Pakstis AJ, Speed WC, Kidd JR, Kidd KK. Candidate SNPs for a universal individual identification panel. Hum Genet. 2007;121(3-4):305-317. doi:10.1007/s00439-007-0342-2.
9. Saremi MA, Saremi M, Tavallaei M. Rapid genomic DNA extraction (RGDE). Forensic Sci Int Genet Suppl Ser. 2008;1(1):63-65. doi:10.1016/j.fsigss.2007.12.001.
10. Quintana-Murci L, Chaix R, Wells RS, et al. Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. Am J Hum Genet. 2004;74(5):827-845. doi:10.1086/383236.
11. 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.
12. 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.
13. Sharafi Farzad M, Tomas C, Borsting C, et al. Analysis of 49 autosomal SNPs in three ethnic groups from Iran: Persians, Lurs and Kurds. Forensic Sci Int Genet. 2013;7(4):471-473. doi:10.1016/j.fsigen.2013.04.001.
14. Heaton MP, Harhay GP, Bennett GL, et al. Selection and use of SNP markers for animal identification and paternity analysis in U.S. beef cattle. Mamm Genome. 2002;13(5):272-281. doi:10.1007/s00335-001-2146-3.
15. Parson W, Ballard D, Budowle B, et al. Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements. Forensic Sci Int Genet. 2016;22:54-63. doi:10.1016/j.fsigen.2016.01.009.
16. de la Puente M, Santos C, Fondevila M, et al. The Global AIMs Nano set: A 31-plex SNaPshot assay of ancestry-informative SNPs. Forensic Sci Int Genet. 2016;22:81-88. doi:10.1016/j.fsigen.2016.01.015.
17. Ambers AD, Churchill JD, King JL, et al. More comprehensive forensic genetic marker analyses for accurate human remains identification using massively parallel DNA sequencing. BMC Genomics. 2016;17(Suppl 9):750. doi:10.1186/s12864-016-3087-2.
18. Bulbul O, Cherni L, Khodjet-El-Khil H, Rajeevan H, Kidd KK. Evaluating a subset of ancestry informative SNPs for discriminating among Southwest Asian and circum-Mediterranean populations. Forensic Sci Int Genet. 2016;23:153-158. doi:10.1016/j.fsigen.2016.04.010.
19. Santos HC, Horimoto AV, Tarazona-Santos E, et al. A minimum set of ancestry informative markers for determining admixture proportions in a mixed American population: the Brazilian set. Eur J Hum Genet. 2016;24(5):725-731. doi:10.1038/ejhg.2015.187.
20. Huang E, Liu C, Zheng J, et al. Genome-wide screen for universal individual identification SNPs based on the HapMap and 1000 Genomes databases. Sci Rep. 2018;8(1):5553. doi:10.1038/s41598-018-23888-0.