Autosomal Short Tandem Repeat (STR) Variation Based on 15 Loci in a Population from the Central Region (Riyadh Province) of Saudi Arabia

Awad E Osman[1]*, Habiba Alsafar[2], Guan K Tay[3], Jasem BJM Theyab[4], Mohamed Mubasher[1], Nezar Eltayeb-El Sheikh[1], Hanan AlHarthi[1], Michael H. Crawford[5] and Gehad El Ghazali[6]

[1]PCLM, King Fahad Medical City, Riyadh 11525, Saudi Arabia
[2]Khalifa University of Science Technology and Research, Abu Dhabi, United Arab Emirates
[3]Centre for Forensic Science, The University of Western Australia, Crawley, Western Australia
[4]Department of Sociology and Social Work and Anthropology, Kuwait University, Kuwait
[5]Laboratory of Biological Anthropology, University of Kansas, USA
[6]Sheikh Khalifa Medical City, PaLMS, SEHA, Abu Dhabi, United Arab Emirates

Abstract

Introduction: The small size of Short Tandem Repeats (STRs), their ubiquitous genome-wide distribution and polymorphic nature enhances their value in human forensic/population genetics applications.

Objectives: This study aims to investigate the short tandem repeat variation based on 15 loci in a population from the central region of Saudi Arabia.

Methods: Allele frequency variation for 15 Short Tandem Repeat (STR) loci was examined in 190 unrelated Saudi volunteers.

Results: This study summarizes the allele distribution in the Saudi population and compares them to other populations located in Asia, Africa, the Middle East and Europe. The standard forensic parameters of Observed Hetrozygosity (Ho), Expected Heterozygosity (He) and Gene Diversity Index (GD) were determined for the following 15 STR loci: D8S1179, D21S1, D7S820, CSF1PO, D3S1358, TH0, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S5, D5S818 and FGA. The most frequent alleles in the Saudi population were: 8 repeats (0.558) at TPOX, 12 (0.411) at D13S317, 12 (0.385) at CSF1PO, 11 (0.382) at D16D539 and 10 (0.358) at D7S820. The 15 markers utilized in this study are highly informative as evidenced by their high power of discrimination (PD) values with D2S1338, D19S433 and FGA having the highest PD values. The relationship between the Saudi population and other geographically distributed populations, assessed by a Multidimensional Scaling (MDS) plot, showed that the Saudi population clustered with groups from Yemen, Iraq, Qatar, Oman and Bahrain.

Conclusion: TPOX, D13S317, CSF1PO, D16D539 and D7S820 markers were found suitable for forensic analysis, paternity testing and can also be used for chimerism study after allogenic bone marrow transplantation for Saudi population. On the other hand, the population admixture with other ethnic origins might explain the variable degree of genetic distances of this population and other Arab-related groups.

Keywords: Short tandem repeat; Saudi arabia; Allele frequencies

Abbreviations: STR(s): Short Tandem Repeat(s); MDS: Multidimensional Scaling; HWE: Hardy–Weinberg Equilibrium

Background

Short Tandem Repeats (STRs) are nucleotides sequences with repeat motifs of variable lengths (2 to 8bp) that are polymorphic (i.e. number of repeats varies between individuals) [1,2]. STRs represent about 3% of human DNA and occur approximately one time in every 10,000 nucleotides. STRs are an invaluable tool and their unique sequences used in genetic finger printing during forensic investigations involve significant biological evidence. Their high degree of polymorphisms makes them informative [3], especially when considering multiple loci simultaneously. The new STR kits now contain over 20 STR loci that are amplified in a single multiplex reaction. Additionally, STR markers have been widely used in medical applications such as for the assessment of allogenic bone marrow transplantation engraftment in calculating the ratio of donor/patient DNA presence and in the study of population genetics. The use of STRs depends on the allele frequencies distributions that vary between various populations [1-4].

Various DNA-based techniques have been used to identify the genetic differences in human populations. STR loci are useful and preferred because of their small size, relatively low incidence of mutation and wide spread distribution [5]. Genetic studies and in particular those based on STR applications have been very important towards developing an appreciation of the extent of genetic polymorphisms that exists between different populations.

Throughout history, society had been stratified on the basis of caste, class, clan, race, region, religion, ethnicity, gender, age and socioeconomic status. It is ethnicity and racial discrimination that distinguishes one nation from the other. Ethnicity is, as defined by Macionis; "a shared cultural heritage and people define themselves or others as members of an ethnic category based on common ancestry, language or religion that gives them a distinctive social identity" [6]. The

*Corresponding author: Awad E Osman, PCLM, King Fahad Medical City, Riyadh 11525, Saudi Arabia, Tel: 966-1-2880000; Fax: 966-1-2881215; E-mail: awadelsid@yahoo.com

same is the case within the Arab world, which has maintained a unique ethnographic identity, historical background, ancestry, cultural traits, social norms, moral values, religious beliefs and genealogy. At present, the total Arab population is estimated at about 325 million, increasing at a rate of approximately 2.3% annually. The Arab population of the Middle East and North Africa is distributed throughout 17 different countries, namely: Algeria, Bahrain, Egypt, Iraq, Jordan, Kuwait, Lebanon, Libya, Morocco, Oman, Palestine, Qatar, Saudi Arabia, Syria, Tunisia, the United Arab Emirates, and Yemen.

In Saudi Arabia, approximately 87% of the population is of Arab descent, a group of people who have inhabited this region for thousands of years. The remaining 13% of inhabitants of this region migrated into this country during the last, 400 years [7]. The rate of consanguineous marriage is relatively high and has been estimated to at around 57.7% [8]. The Saudi population can be characterized by a specific ethnic subset which presumably reflects a unique set of allele frequencies and genotypes and that could differentiate them from other populations in near geographical areas and worldwide i.e. the alleles frequencies will be different from other populations and the genetic distances between these populations will be variable depending on the time of divergence from the common human ancestor.

Therefore, this study was conducted to investigate the genetic variation of populations from Saudi Arabia by using 15 autosomal STR markers - currently used in forensic and paternity testing. DNA was collected from a sample of 190 unrelated volunteers residing in the central region (Riyadh Province) of Saudi Arabia. The allele frequencies of the 15 STR markers were calculated using gene counting methods and compared with frequencies from other populations previously studied.

Methods

Population

A total of 190 healthy unrelated individuals from the central region (Riyadh Province) of Saudi Arabia were randomly recruited at King Fahad Medical City (KFMC) (Riyadh, Saudi Arabia) from the donors of patients under preparation for Bone Marrow Transplantation. According to KFMC hospital record system, the selection of the study subjects precluded having any two individuals or their father/mother descending from the same parent. Thus, first and second cousins are not included in the study. The age of the study population ranged between 10 and 45 years (median of 25) at the time of blood collection. There were 131 (68.9 %) males and all were of Saudi origin. The sample size was selected to provide sufficient analytical power in terms of degrees of freedom to determine allele frequencies that will allow testing of the Hardy Weinberg Equilibrium (HWE) assumption based on a Fisher's Exact Test. The study received ethical approval from the Institutional Review Board at KFMC.

DNA extraction, PCR and Fragment analysis for STR markers

Genomic DNA was extracted from whole blood samples in EDTA anticoagulant using a MagNa pure compact instrument (Roche Diagnostics GmbH, Mannheim, Germany). PCR amplification was performed according to manufacturer instruction on GeneAmp® PCR System **9700 Thermal Cycler** (Applied Biosystems, Foster City CA, USA) using AmpFlSTR® Identifiler® PCR amplification kit (Applied Biosystems, Foster City CA, USA) that include D8S1179, D21S1, D7S820, CSF1PO, D3S1358, THO, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S5, D5S818, FGA loci. The PCR product was separated on the 3130xL Genetic Analyzer (Applied Biosystems,

Foster City, CA, USA) and analyzed using GeneMapper® Software v3.2 (Applied Biosystems, Foster City CA, USA). For quality assurance purposes, multiple DNA samples with known combined STR genotypes were run in parallel with our study samples and appropriate results were obtained for these control samples on every occasion.

Statistical analysis

The allele frequencies were calculated based on the number of the detected alleles for each specific locus. To assess departures from Hardy-Weinberg Equilibrium (HWE) (observed heterozygosity (Ho), expected heterozygosity (He), p-value; using Fisher's Exact test) and Gene Diversity indices (GD), the Arlequin software version 3.5.1.3 was used [9]. Forensic and population genetic parameters including power of discrimination (PD), Polymorphic Information Content (PIC), Matching probability (MP) and probability of exclusion (PE) were derived utilizing Powerstats software version 1.2 [10]. An overall significance level for testing HWE hypothesis was set at 0.05 and adjusted for multiple comparisons using Bonferonni Criteria. Genetic relationships between groups were assessed through multidimensional scaling (MDS) plot of the genetic distance matrix using NTSYS 2.1. Software to determine the genetic differences and similarities between this study sample and published data from other populations [11].

Results and Discussion

The 15 autosomal STR loci from all samples tested were amplified successfully. A total of 150 alleles were identified for the 15 STR loci. Their corresponding frequencies were calculated (Table 1). The highest allelic frequencies observed were: allele 10 of TPOX (55.8%) and allele 12 of D13S317 (41.1%) with the most polymorphic loci observed in this study being D18S5, FGA, D21S11--defined by 17, 15 and 15 alleles respectively. However, the Polymorphic Information Content (PIC) of the 15 loci were greater than 0.57 through all loci that suggesting that the markers are highly polymorphic and would be useful as informative markers for differentiating individuals of Saudi descent. The degree of polymorphism at each locus can also be expressed in terms of heterozygosity along with the PIC value [12]. The highest values were observed for the D19S433 locus (He=0.86977; PIC=0.85) while the lowest values were identified in TPOX (He=0.6212; PIC=0.57).

After applying the Bonferonni corrections to the Fisher's exact test a deviation from Hardy–Weinberg equilibrium (HWE) was detected for the TH0, D5S818 and FGA loci. A deviation from HWE was also detected in a previous Saudi study of HLA allele frequencies [13]. This deviation was attributed to the high rates of consanguineous marriage among Saudis or to the Wahlund effect resulting in the reduction of heterozygosity. Findings in a previous Saudi study observed that 2 out of 8 STR loci did not conform to HWE which is consistent with our data [14]. As shown in Table 2, the p values for HWE of the available STR loci data that were obtained from previous Arab related populations reports were compared with the data from this study. It is worth noting that some alleles such as 34.2 (D21S11), 16.2 (D18S51), 12.2 (D16S539), 29 and 22.2 (FGA) and 7.3 and 8.3 (THO1) that were considered to be variant alleles in previous reports were also identified in frequencies ranged between 0.3 to 1.6% within the Saudi population studied in this project.

This study was based on a random sample of unrelated healthy volunteers of Arab descent who reside in the Riyadh district of Saudi Arabia. Allelic frequencies of 15 STR markers and their population genetic/forensic parameters were analyzed and results showed that the polymorphic nature of loci examined were sufficient to allow the

allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
5.3						0.032									
6				0.004		0.284									
7			0.016			0.195									
7.3						0.003									
8			0.134			0.15	0.113	0.008				0.558		0.008	
8.3						0.003									
9	0.008		0.137	0.008		0.232	0.053	0.158				0.187		0.068	0.003
9.3						0.097									
10	0.063		0.358	0.294		0.005	0.029	0.071				0.092	0.005	0.087	
10.2													0.003		
11	0.084		0.187	0.246			0.182	0.382		0.021		0.161	0.028	0.226	
12	0.132		0.134	0.385			0.411	0.229		0.095		0.003	0.188	0.384	
12.2									0.003	0.003					
13	0.208		0.021	0.029			0.155	0.124	0.003	0.161	0		0.268	0.195	
13.2										0.061					
14	0.244		0.003	0.008	0.045		0.058	0.026	0.008	0.205	0.08		0.108	0.026	
14.2										0.076					
15	0.205				0.192					0.121	0.2		0.149	0.005	
15.2										0.166					
16	0.066				0.305				0.068	0.032	0.25		0.095		
16.2										0.05			0.005		
17	0.008				0.289				0.216	0.003	0.24		0.055		
17.2										0.008					
18					0.129				0.092		0.16		0.045		0.003
19	0.003				0.32				0.164		0.07		0.016		0.053
20					0.008				0.198		0		0.008		0.082
21									0.038						0.213
21.2															0.003
22									0.027				0.011		0.116
22.2															0.003
23									0.047				0.005		0.134
24									0.078				0.008		0.187
24.2															0.003
25									0.047				0.003		0.105
26									0.003						0.079
27		0.013													0.003
28		0.105													
29		0.289													0.016
29.2		0.005													
30		0.211													
30.2		0.008													
31		0.079													
31.2		0.074													
32		0.011													
32.2		0.111													
33		0.008													
33.2		0.066													
34.2		0.008													
36		0.003													
38		0.008													
Ho	0.9	0.815	0.758	0.6705	0.7368	0.774	0.7474	0.7579	0.7869	0.8053	0.72	0.621	0.845	0.653	0.811
He	0.834	0.834	0.782	0.713	0.7685	0.797	0.7566	0.7579	0.8596	0.8698	0.81	0.621	0.844	0.752	0.863
p-value	0.0739	0.004	0.457	0.23	0.0827	0.0002	0.0933	0.0279	0.0037	0	0.08	0.407	0.005	0.0008	0.0003
GD	0.834	0.829	0.763	0.6835	0.7684	0.797	0.7566	0.7579	0.8336	0.8698	0.81	0.621	0.832	0.752	0.864
MP	0.061	0.057	0.082	0.136	0.095	0.079	0.097	0.102	0.042	0.041	0.07	0.194	0.048	0.098	0.041
PD	0.939	0.943	0.918	0.864	0.905	0.921	0.903	0.898	0.958	0.959	0.93	0.806	0.952	0.902	0.959
PIC	0.81	0.81	0.75	0.66	0.73	0.76	0.72	0.72	0.84	0.85	0.78	0.57	0.82	0.71	0.85
PE	0.795	0.627	0.524	0.384	0.488	0.551	0.505	0.523	0.575	0.609	0.46	0.317	0.685	0.359	0.619

Table 1: Allele Frequency Distribution for the 15 STR loci in a Sample of Saudi population (n=190)

Population	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
Saudis (current)	0.0739	**0.0036**	0.4573	0.23	0.0827	**0.0003**	0.0933	**0.0279**	**0.0037**	**0**	0.0806	0.4069	**0.005**	**0.0008**	**0.0004**
Saudis [9]	N/A	**N/A**	N/A	0.037	N/A	**0.094**	N/A	**N/A**	**N/A**	**N/A**	0.858	0.821	**N/A**	**N/A**	**N/A**
Tunisians [4]	**0.0134**	**0.0414**	N/A	N/A	0.0764	0.3144	N/A	N/A	0.1897	**0**	**0.009**	N/A	**0.001**	N/A	**0.0116**
Jordanians [14]	0.997	0.997	0.75	0.874	0.277	0.421	0.066	0.108	N/A	N/A	0.673	0.952	0.275	0.471	**0.004**
Palestinians [7]	0.865	0.424	0.794	0.173	0.141	0.157	**0.019**	0.081	N/A	N/A	**0.006**	0.474	0.455	**0.009**	0.285
Iraqis [12]	0.06	0.135	0.393	0.533	0.824	0.936	0.66	0.25	0.473	0.344	0.058	0.377	0.399	0.45	0.71
Omanis [13]	0.282	0.333	0.32	N/A	0.56	N/A	0.082	N/A	N/A	N/A	0.254	N/A	0.204	0.644	0.533
Egyptians [11]	0.139	0.598	0.527	N/A	0.695	N/A	0.08	N/A	N/A	N/A	**0.031**	N/A	0.067	0.52	0.223
Moroccans [15]	0.144	N/A	0.441	0.692	0.304	0.111	0.109	0.903	N/A	N/A	**0.002**	0.449	N/A	0.414	0.848
Syrians [15]	0.781	0.094	0.423	0.712	0.349	0.703	0.788	0.74	N/A	N/A	0.735	0.242	0.775	0.89	0.6

Table 2: Comparison p values of HWE for STRs data for nine Arab related populations. *Bold indicates a significant difference between the observed and the expected heterozygosity.

characterization of individuals from Saudi Arabia (markers D21S1, THO, D2S1338, D19S433, D18S5, D5S818 and FGA). This study showed that D2S1338, D19S433 and FGA were the most polymorphic markers as evidenced by their high PD values. It is worth noting that the 4 markers with high PD values (vWA, THOI, CSFIPO and TPOX) were also identified by a previous study that used 8 STR loci to study unrelated subjects in Saudi Arabia (14).

In other studies, the most frequent allele for forensic STR loci can vary depending on the population, for example 12 (0.4093) at D5S818 in Slovenian population [15], 15 (0.4143) at D3S1358 in Bolivians [16], 8 (0.4890) at TPOX among Wallachians in South Romania [17] and 8 (0.424) at TPOX in Adaima community from Egypt [18]. In this study the most frequent alleles were 8 (0.558) at TPOX, 12 (0.411) at D13S317, 12 (0.385) at CSF1PO, 11(0.382) at D16D539 and 10 (0.358) at D7S820. The most polymorphic marker in this study was D18S51; spanning 17 tandem repeat alleles. Allele 13 of this locus was the most predominant one with a frequency of 0.268. Interestingly, a study of Tunisians found markers D19S433 and D21S11 to be the most polymorphic; each spanning 18 alleles [4]. In contrast, a lesser degree of polymorphism was reported for Iraqi individuals as compared to Saudis and Tunisians [19] due to less admixture.

For the sample used in this study, the observed values of heterozygosity ranged from 0.621 at TPOX to 0.869 at D19S433. This lower degree of homozygosity suggests presence of a random mating in the study population and less consanguinity. Despite prolific rates of consanguineous marriage in the Saudi population [8], this study was able to clearly identify seven markers that can be used for characterizing the genetic makeup among individuals of Arab descent in Saudi Arabia.

Furthermore, it can also be argued that Independent Inheritance property could apply to two cases for markers collocated on the same chromosome, namely CSF1PO and D5S818 on chromosome 5, and TPOX and D2S1338 on chromosome 2. This was reflected by the disparate statistical values regarding HWE assumption that rendered in each case one marker as statistically significant while the other one was not.

It is also interesting to note that variant alleles of 34.2 (D21S11), 16.2 (D18S51), 12.2 (D16S539), 29 and 22.2 (FGA) as well as 7.3 and 8.3 (THO1) have also been identified by other studies. The 15 markers studied here were highly informative as illustrated by their high PD values (Table 1). Consistently higher PD values were also reported studies that examined populations from South Romania at locus FGA (PD=0.970), Tunisian at locus D19S433 (PD=0.976), sub-population of Botswana at locus D21S11 (PD=0.974) and Turkish at locus D18S51 (PD=0.979) [3,4,17,20].

A Multidimensional Scaling (MDS) plot was constructed to illustrate the genetic distances between 14 populations (Figure 1). The analysis of these populations showed clustering into four groups: (1) the Asian subpopulation, (2) the Caucasians, (3) African and (4) Middle East subpopulations. The Saudi group in this study clustered with populations from Yemen, Iraq, Qatar, Oman and Bahrain. In contrast, African populations like the Hutu and Kenyans as well as the Pakistani, Bangladeshi, Punjabi and those of Caucasian origin like the Belgium, Croatians and Georgians were located furthest from the samples of this study.

Conclusions

TPOX, D13S317, CSF1PO, D16D539 and D7S820 markers were found suitable for forensic analysis, paternity testing and can also be used for chimerism study after allogenic bone marrow transplantation for Saudi population. On the other hand, the variable degree of genetic distances of this population and other Arab related groups might be explained by the populations admixture with other ethnic origins. Further analyses of Arab population are needed to understand the interrelationship between the different ethnic groups of the region.
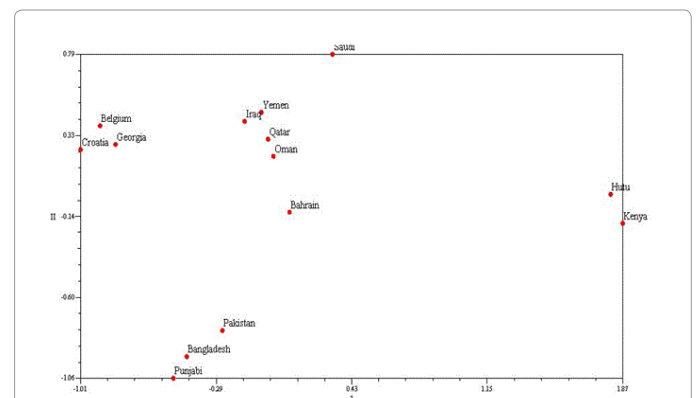
Figure 1: Multidimensional Scaling plot of genetic distance between 14 populations: Saudi Arabia, Yemen, Iraq, Qatar, Oman, Bahrain, Pakistan, Bangladesh, Punjabi, Belgium, Croatia, Georgia, Hutu and Kenya. The Saudi population studied grouped with other populations from the Middle East.

References

1. Krstic ASO, Guc-Scekic M, Vujic D and Varljen T (2007) Hematopoietic stem cell transplantation monitoring in childhood. Hematological diseases in Serbia: STR-PCR techniques. Arch biol Sci 59: 23-27.

2. Zhang W, Yu YY (2007) Polymorphisms of short tandem repeat of genes and breast cancer susceptibility. Eur J Surg Oncol 33: 529-534.

3. Mpoloka SWK (2008) Determination of allele frequencies in nine short tandem repeat loci of five human sub-populations in Botswana. Afr J Biotech 7: 1023-1027.

4. Souiden Y, Chaieb K, Chehab O, Mahdouani K (2009) Allele frequencies of ten short tandem repeats loci in the central Tunisian human population. J Genet 88: 119-122.

5. Kashyap VSN, Sahoo S, Sarkar B and Trivedi R (2003) Genetic variation at fifteen microsatellite loci in human populations of India. Curr science 85: 464-473.

6. Macionis J (2007) Sociology. 11 ed. New Jersey: Pearson.

7. Lindsey JE (2005) Daily Life in the Medieval Islamic World. Indianapolis: Hackett Publishing Company Inc.

8. el-Hazmi MA, al-Swailem AR, Warsy AS, al-Swailem AM, Sulaimani R, et al. (1995) Consanguinity among the Saudi Arabian population. J Med Genet 32: 623-626.

9. Excoffier L, Laval G, Schneider S (2007) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online 1: 47-50.

10. A T. Tools for analysis of Population statistics. Profiles DNA 1999:2: 14-16, Powerstats version 1.2, Promega Corporation.

11. FJ. R. NTSYS-PC: Numerical taxonomy and multivariate analysis system, 2.1 ed. New York (NY): Exeter Software. 2005.

12. Abu Halima MS, Bernal LP, Sharif FA (2009) Genetic variation of 15 autosomal short tandem repeat (STR) loci in the Palestinian population of Gaza Strip. Leg Med (Tokyo) 11: 203-204.

13. Osman A, Mubasher M, Al-Harthi H, ElSheikh N, Elghazali G (2014) HLA-A, -B, -C, -DRB, and -DQB1 Allele Lineages and Haplotype Frequencies among Saudis. Immunol Immunogen Insights 6: 1- 6.

14. Sinha S, Amjad M, Rogers C, Hamby JE, Tahir UA, et al. (1999) Typing of eight short tandem repeat (STR) loci in a Saudi Arabian population. Forensic Sci Int 104: 143-146.

15. Drobnic K PN, Bakal N and Marjanovic D (2005) Allele frequencies for the 15 short tandem repeat loci in Slovenian population. J Forensic sci 50: 1-3.

16. Cifuentes L, Jorquera H, Acuña M, Ordóñez J, Sierra AL (2008) Allele frequencies for 12 autosomal short tandem repeat loci in two Bolivian populations. Genet Mol Res 7: 271-275.

17. Stanciu F, Stoian IM, Popescu OR (2009) Population data for 15 short tandem repeat loci from Wallachia Region, South Romania. Croat Med J 50: 321-325.

18. Coudray C, Guitard E, el-Chennawi F, Larrouy G, Dugoujon JM (2007) Allele frequencies of 15 short tandem repeats (STRs) in three Egyptian populations of different ethnic groups. Forensic Sci Int 169: 260-265.

19. Barni F, Berti A, Pianese A, Boccellino A, Miller MP, et al. (2007) Allele frequencies of 15 autosomal STR loci in the Iraq population with comparisons to other populations from the middle-eastern region. Forensic Sci Int 167: 87-92.

20. Ulküer U, Kurtuluş-Ulküer M, Elma C, Kesici T, et al. (2004) Short tandem repeat (STR) polymorphisms in Turkish population. J Genet 83: 197-199.