Comparative analysis of Y-chromosomal short tandem repeats (YSTRs) polymorphism in an Iranian Sadat subpopulation

Farazmand A. [1]*, **Sokhansanj A.** [1], **Rafee M.R.** [1], **Mehrabani Yeganeh**[2]
[1] *Department of Cell and Molecular Biology, School of Biology, University of Tehran, Tehran, Iran.*
[2] *Department of Animal Sciences, University College of Agriculture, University of Tehran, Tehran, Iran.*
[*] *Corresponding author, e-mail: afarazmand@khayam.ut.ac.ir*
(*received: 17/5/2008 ; accepted: 19/7/2009*)

Abstract

Short and highly polymorphic DNA markers, so called short tandem repeats (STR), have become an important and efficient means in anthropology and forensic science. The Y chromosome STRs are male specific markers reside outside of pseudoautosomal region and path trough male lineage. Moreover the Y-specific markers may found to be associated with Y-linked phenotypes, including male infertility and gender-specific cancers. In current study we have examined YSTR polymorphism in a random sample of males among Iranian Sadat subpopulation. Sadat refers to descendants of Prophet Mohammad (through paternal lineage common to the first Imam of Shiite and his cousin Prophet Mohammad). The choice of this group had the advantage of being a decent lineage related through paternal lineage traced back to the common ancestor according to respective pedigrees. We have analyzed a random sample of the Sadat population using 6 Y-STR markers (DYS19, DYS385a, DYS389I, DYS390, DYS392, and DYS393). Following the total DNA extraction from blood sample, PCR is performed and the products are visualized on agarose gel followed by denaturing polyacrylamid gel electrophoresis for further resolution. Comparing our date to that of some related data on other Iranians, Arab and Turkish male population revealed a significant lower genetic diversity in DYS385a, DYS390, and DYS392, and most notably at the latter locus.

Keywords: Y chromosome, Sadat, STRs polymorphism, Population Genetics.

1- Introduction

Short tandem repeats (STRs), commonly referred to as microsatellites, are loci characterized by the tandem repeated arrays of 2-6 bp core units. In human, they are estimated to make up about 2% of the genome. The basis of the STRs power as markers for discrimination is length polymorphism that is common in human population. The polymorphic nature of STRs, ranging between 100 – 400 bp, and their common presence on all 22 autosomal chromosomes and on both sex chromosomes, and to be amenable to degraded DNA analysis due to the short length of their tandem repeated core units make them highly suitable for anthropological and archeological and forensic applications (Iida & Kishi, 2005, Gusmao *et al.* 2005).

In contrast to autosomal STRs, the Y-chromosomal STRs represent a haplotype, as they have no homologue counterpart on X chromosome. The Y chromosome passes from father to the son. The vast majority of the time, father passes an exact copy of his Y-chromosome to his son. This means that the Y chromosome linked markers of the son are identical to those of his father. However, rarely there may be a mutation. The average mutation rate for Y-STR loci is estimated to be $3.17 * 10^{-3}$, although the rate varies among different markers ranged from 0 to $8.85 * 10^{-3}$ (Kayser *et al.* 2000). Due to relatively high mutation rate, the Y-STR provides a forensic tool for specific problems, such as identification of crime suspects and paternity testing. Also Y-chromosomal STR typing enables the follow-up of paternal lineages. Consequently, Y-chromosome haplotyping may be an interesting additional tool for certain cases in the reconstruction of family histories.

Sadat refers to a population whose members are descendents of Imam Ali (the cousin of Prophet Mohammad and the first imam of Shiite). Originally they lived in Iraq and Saudi Arabia regions but gradually some of them migrated to Iran. Y-STR analysis of Sadat has the advantage that most of them show a pedigree which reveals their paternal lineage through Imam Ali. Also, the Sadat title is given to ones who were offspring of a Sadat man.

Obviously there are mutations that have occurred during last 14 centuries of Sadat history. Searching

for these mutations and generally finding possible differences and similarities among a male population of Sadat compared to a population of the Iranians living in Tehran and Isfahan (Nasidze *et al.* 2003), male population of Arabs of Syria (Abdin *et al.* 2003), and Turkish of Central Anatolia (Hadi Cakir *et al.* 2004) was the purpose of this study.

We therefore analyzed 50 unrelated volunteer putative Sadat according to their pedigrees, by Y-chromosomal STR loci. The gene diversity, average gene diversity and haplotype diversity of the samples were calculated in Sadat and compared with above mentioned populations.

2- Material and methods

2-1- Target population

Total numbers of 50 males were selected from unrelated Sadat males, chosen from different cities of Iran, to eliminate the possibility of testing related males (Tab. 1). Also target population was checked for pedigree declaring their paternal lineage through Imam Ali.

Table 1. Different cities of Iran where volunteers belong to.

City	*Sample #*	*City*	*Samples #*
Amol	5	Kermanshah	1
Ardehal	1	Khansar	2
Babol	3	Khoramabaa	1
Babolsar	1	Langrood	1
Bojnoord	1	Mashhad	2
Eshkevar	1	Ormia	2
Fuman	1	Sabzevar	1
Gaznein	1	Shiraz	2
Gha'en	1	Shooshtar	5
Ghamsar	1	Some'esara	1
Ghom	1	Tabriz	3
Gorgan	1	Tehran	1
Goulpayga	1	Yasooj	3
Isfa han	2	Yazd	3
Kashan	1		

2-2- DNA extraction

Blood samples were obtained, and genomic DNA was extracted by Phenol-Chloroform (Sambrook J. et al., 1989) procedure. Quantification and quality of recovered DNA was determined by spectrophotometeric analysis (Sambrook *et al.* 1989). Two micro-liter (between 10-20ng) of genomic DNA was used for each PCR reactions.

2-3- Amplification conditions

Six loci (seven polymorphic markers) including DYS19, DYS385, DYS389II, DYS390, DYS392 and DYS393 were amplified in this study. The respective primer sequences for the loci were chosen according to Butler (2002) and the characteristics of the loci and primer sequences are given in Tab. 2. Moreover PCR was optimized for DYS385-DYS390 and DYS393-DYS19 to be amplified in two duplex reactions and another two monoplex for DYS392 and DYS389II.

Further PCR conditions were 2.5 units Taq DNA polymerase, 200µM dNTP with 2mM $MgCl_2$ and primer concentration of 0.20µM for DYS385, DYS389, DYS393, DYS19 and 0.40µM for DYS390 in a total volume of 50µl. Cycling conditions were as follows: Hot start at 95°C for 7 min followed by 30 cycles of 94°C 30 s, 56°C 1 min, 72 °C 1 min and a final extension step at 72 °C for 10 min. Also annealing temperature was raised up to 59 °C for DYS390-DYS385 duplex reaction.

2-4- Gel electrophoresis condition

First all PCR products were visualized by agarose gel electrophoresis (Fig. 1), then length polymorphism of the fragments was observed on 6% acrylamide/bisacrylamid (19:1) denatured (8M urea) slab gel electrophoresis 1200 V, 30W, 50CM separation distance. Bands were stained based on the silver staining protocol (Qu *et al.* 2005; Fig. 2). Visualized bands were compared to a self-constructed sequenced ladder.

2-5- Statistical analysis

According to Nei's suggestion (Nei 1987) genetic diversity (GD) was calculated using the following formula:

$$GD = \frac{n(\sum(1-x^2)}{(n-1)}$$

Where n is the number of the individuals, and x represents the allelic frequency in a given population sample. In Y-linked polymorphism, haplotype diversity is numerically identical to the power of discrimination (PD) and the chance of exclusion (CE) (Abdin *et al.* 2003).

Standard error (S.E.) was calculated as:

$$S.E. = \{\frac{2}{n}[\sum x^3 - (\sum x^2)^2]\}^{1/2}$$

Furthermore, all of these parameters were also calculated using Arlequin software (Schneider *et al.* 2000).

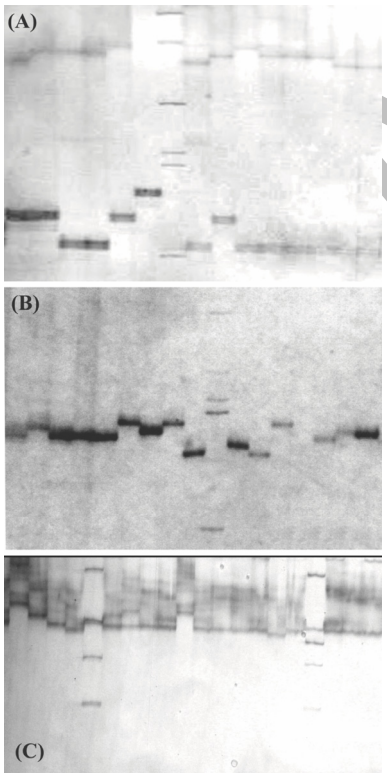
3- Results and Discussion

3-1- Multiallelic pattern

More than one allele per locus (and more than two alleles at DYS385 marker allele) in a single source

Table 2. Observed allele/genotype frequencies of the 7 Y-STR loci in Sadat of Iran.

Alleles	*DYS19*	*DYS389II*	*DYS390*	*DYS392*	*DYS393*	*DYS385a*	*DYS385b*	*Genotype*	*DYS385a/b*
10				0.1000		0.0400		10-13	0.0400
11				0.8000		0.0800		11-12	0.0200
12				0.0600	0.4313	0.2800	0.1200	11-13	0.0200
13	0.1000			0.0200	0.2941	0.4600	0.1400	11-14	0.0200
14	0.4800			0.0200	0.2549	0.0600	0.0400	11-20	0.0200
15	0.2600				0.0000	0.0400	0.2800	12-12	0.1000
16	0.1600				0.0196	0.0400	0.1200	12-13	0.0400
17							0.1400	12-15	0.1200
18							0.0800	12-17	0.0200
19							0.0600	13-13	0.0400
20							0.0200	13-14	0.0200
21			0.0833					13-15	0.1400
22			0.0625					13-16	0.0600
23			0.5208					13-17	0.1000
24			0.2083					13-18	0.0600
25			0.1250					13-19	0.0400
26								14-16	0.0200
27								14-17	0.0200
28		0.0200						14-19	0.0200
29		0.4000						15-15	0.0400
30		0.3800						16-16	0.0200
31		0.2000						16-17	0.0200
32	0.6800	0.6686	0.6729	0.3527	0.6753	0.7094	0.8588		0.9457
33									
S.E.	0.0436	0.0285	0.0565	0.0827	0.0282	0.0481	0.0250		0.0106
GD1	0.6706	0.6478	0.7317	0.5640	0.6338	0.8038	0.8540		0.9017
GD2	0.6642	0.7489	0.6361	0.4527	0.7159	0.7682	0.8016		0.9534
GD3	0.6954	0.7376	0.7161	0.5994	0.6596	0.7959	0.8268		0.9532



Firgure 1. Agarose gel electrophoresis Duplex PCR of DYS19, DYS393 (A), duplex PCR of DYS385, DYS390 (B), monoplex PCR of DYS389II and DYS392 (C). Ladder bands show 50-100-150-200-250-300-400-500-600-700-800-900-1000 base pair length.

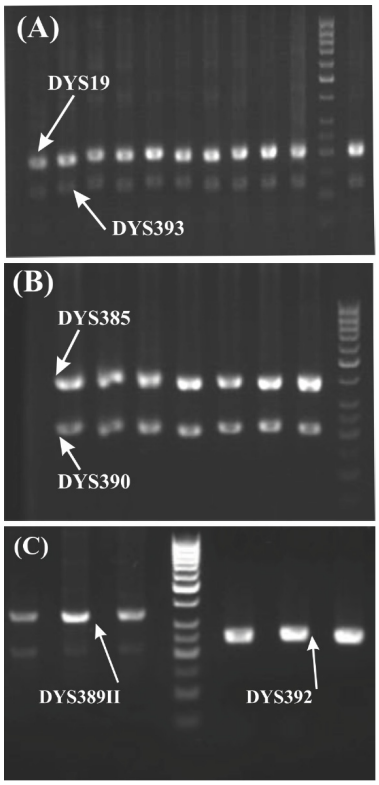


Figure 2. Denaturing polyacrylamid gel electrophoresis. DYS19 and DYS393 (A), DYS390 (B), DYS392 (C).

sample has been reported previously (Santos *et al.* 1996, Kayser *et al.* 2001) Biallelic instances were observed in loci DYS390 and DYS393 in our samples and there was shown a triple allele in DYS390 (Fig. 3). The most plausible explanation to this phenomenon is gene duplication. There was a relatively high frequency of biallelic instances among Sadat.

3-2- Gene diversity and the number of observed alleles

Significantly, Sadat demonstrated a lower genetic diversity in DYS385a, DYS392 and DYS390 compared to that of other Iranian, Arab, and Turkish male populations, most notably at DYS392 loci (Tab. 3).

In Sadat, the most common alleles in each locus were as follow: allele 14 in DYS19; 29 in DYS389II; 23 in DYS390; 11 in DYS392; 12 in DYS393; 13 in DYS385a and 15 in DYS385b. While the most frequent alleles in Iranian, Arab and DYS392; 12, 12, 12 in DYS393; 13, 13, 11 in DYS385a and 14, 16, 16 in DYS385b, respectively. Allelic frequency diagram of the alleles (generated using SPSS software) is illustrated in Figure 4.

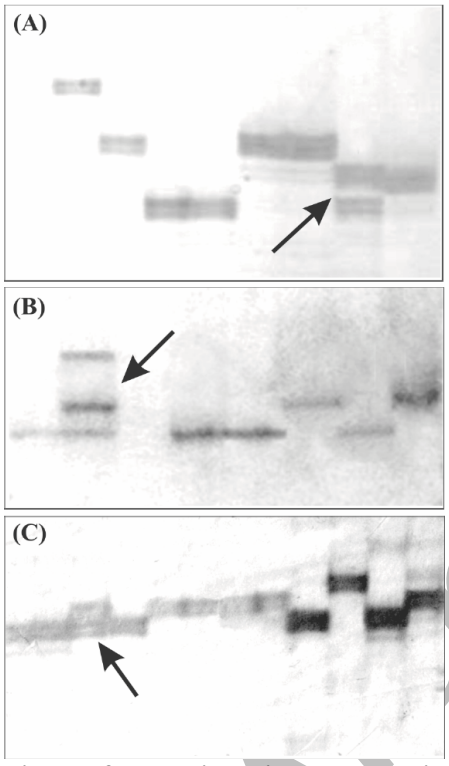


Figure 3. Multi-allelic patterns in Sadat population. A biallelic DYS393 consist of 12 and 13 repeats (A), a tri-allelic DYS390 consist of 23, 24 and 26 repeats (B), a biallelic DYS390 consist of 23 and 24 repeats(C).

Table 3. Distribution of 7 Y-STR loci haplotypes among Sadat of Iran.

Samples	DYS19	DYS385a	DYS385b	DYS389II	DYS390	DYS392	DYS393
H01	16	12	15	30	23	11	14
H02	16	12	15	30	23	11	14
H03	16	12	15	30	23	11	14
H04	15	12	15	30	24	11	14
H05	15	12	15	30	23	11	14
H06	14	13	17	30	24	11	12
H07	14	14	16	31	26,24,23	13	13
H08	14	13	18	30	24	11	12
H09	15	12	12	31	23	11	13
H10	13	11	20	30	23	11	12
H11	15	12	12	31	24	11	13
H12	15	12	12	31	23	11	13
H13	15	12	12	31	24,23	11	13
H14	16	12	12	31	24	11	13
H15	13	11	12	29	23	12	13
H16	14	13	16	29	23	11	12
H17	15	13	15	29	21	11	14
H18	14	13	18	29	23	11	14
H19	14	12	13	29	25,23	11	13,12
H20	14	13	17	30	23	11	12
H21	14	13	15	29	23	11	12
H22	16	15	15	30	22	10	14
H23	15	15	15	30	25	11	13
H24	16	13	15	29	21	11	14
H25	14	13	17	31	22	11	12
H26	15	13	15	29	21	11	14
H27	13	13	14	29	25	13	13
H28	13	16	17	30	24	11	12
H29	14	14	17	29	25	11	12
H30	14	12	17	29	23	11	12
H31	14	13	17	30	23	11	12
H32	14	13	13	29	22	11	12
H33	16	10	13	30	23	11	13
H34	15	12	13	29	21	11	16
H35	15	13	13	30	23	11	12
H36	15	11	14	30	25	11	13
H37	14	13	16	29	23	11	12
H38	14	13	19	29	23	11	12
H39	14	13	18	29	23	11	13
H40	16	16	16	29	22	10	12
H41	14	13	15	31	23	11	13
H42	14	12	15	29	24	12	14
H43	14	14	19	30	24	11	14
H44	14	10	13	31	23	14	13
H45	14	13	15	29	23	11	12
H46	14	13	17	30	23	11	13
H47	15	13	16	30	23	11	13
H48	14	11	13	28	24	12	14
H49	14	13	19	31	23	10	13
H50	15	13	15	29	25	11	15

Turkish males were 14, 14, 14 in DYS19; 29, 30, 30 in DYS389II; 24, 23, 23 in DYS390; 11, 11, 11 in

The average gene diversity by these 7 markers was 0.6564 in Sadat compared to 0.7004, 0.6839 and 0.7187 among Iranian, Arabic and Turkish males, respectively.

For the comparison of pattern of allele distribution in each locus a Pearson Correlation has been made among the population samples. The highest significant correlations were observed at the 0.01 levels (2-tailed) between Sadat and Arab males in allele frequency of DYS390 and DYS385a/b, and also between Sadat and Turkish males in allele frequency of DYS393 and DYS19 and between Sadat and Iranian males in allele frequency of DYS392. DYS389II was significantly (P value> 0.999) correlated between Sadat and all other three nations.

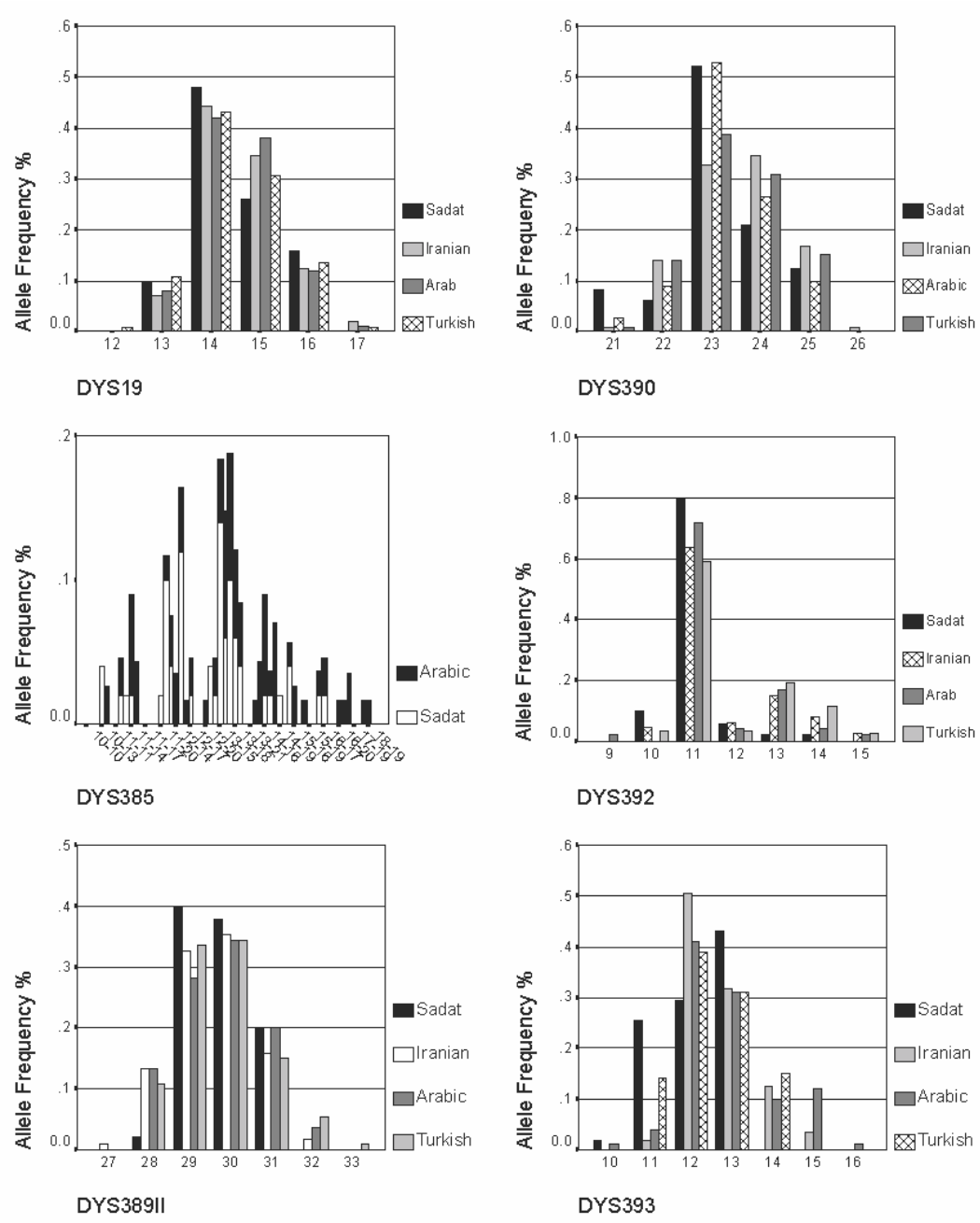


Figure 4. Clustered Bar graph of allele frequencies in regard to the nations. The most similar nation to Sadat in the allele distribution of each locus is illustrated by a crosshatched fill pattern. Pattern of allele distribution of DYS389 is correlated among Sadat and all other three nations. Since DYS385 has many alleles, its bar graph is illustrated in regard to Arab and Sadat, which have the most similar patterns of distributions.

Haplotypes constructs of the six Y-STR loci are shown in Tab. 4. Haplotype Diversity (HD) calculated by using haplotype frequencies of these 7 markers in 50 unrelated Sadat males, was 0.9975, in which 47 unique haplotypes were found among 50 samples; and 3 haplotype were found to be multi-allelic. Similar to this HD value was obtained by 9 markers in Iranian living in Tehran and Isfahan (0.997) (Nasidze *et al.* 2003), also by 13 markers in Arab of Syria (0.9902) (Abdin *et al.* 2003) and by 11 markers in Turkish of Central Anatolia (0.9987) (Hadi Cakir *et al.* 2004).

Although Sadat revealed significant reduction in diversity of some genes, based on the almost highest HD value obtained (0.9975), our findings indicate that these 7 Y-STR markers may be applied for forensics and paternity testing in Sadat populations, where local founder effects have created a substructure.

References

Abdin L., Dewa K., Rand S., Hohoff C., Brinkmann B. 2003: Analysis of 13 Y-chromosomal STRs in an Arab population sample from Syria. *International Congress Series.* **1239:** 319-321.

Butler B., Schoske R., Vallone P. M., Kline M. C., Redd A. J., Hammer M. F. 2002: A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers Forensic Science International. **129:** 10-24.

Gusmao L. Gusmão L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, Morling N, Prinz M, Roewer L, Tyler-Smith C, Schneider PM; DNA Commission of the International Society of Forensic Genetics.. 2006: DNA commission of International Society of Forensic Genetics (ISFG): An update of recommendations on the use of Y-STRs in forensic analysis. *Forensic Science International.* **10:** 187-97.

Hadi Cakir A. H., Celebioğlu A, Yardimci E. 2004: Y-STR haplotypes in Central Anatolia region of Turkey. *Forensic Science International.* **144 :** 59-64.

Iida I.R., Kishi K. 2005: Identification, Characterization and forensic application of novel Y-STRs. *Legal Medicine.* **7:** 255-258.

Kayser M., Roewer L., Hedman M., Henke L., Henke J., Brauer S., Krüger C., Krawczak M., Nagy M., Dobosz T., Szibor R., de Knijff P., Stoneking M., Sajantila A. 2000: Characteristics and frequency of germ line mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs, *Am. J. Hum. Genet.* **66:** 1580-1588.

Kayser M., Sajantila A. 2001: Mutations at Y-STR loci: implications for paternity testing and forensic analysis, *Forensic Science. Int.15;* **118 :** 116-121.

Nasidze I., Schädlich H., Stoneking M.. 2003: Haplotypes from the Caucasus, Turkey and Iran for nine Y-STR loci. *Forensic Science International.* **137:** 85-93.

Nei M. 1987: Molecular Evolutionary Genetics, Columbia University Press. New York. 176-181.

Qu L., Wu G., Yang N. 2005: Efficient and sensitive method of DNA silver staining in polyacrylamide gels. *Electrophoresis.* **26 :** 99-101.

Sambrook J., Fritsch E. F., Maniatis T. 1989: *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor. New York.

Santos F.R., Gerelsaikhan T., Munkhtuja B., Oyunsuren T., Epplen J.T., Pena S.D.. 1996: Geographic differences in the allele frequencies of the human Y-linked tetranucleotide polymorphism DYS19, *Hum. Genet.* **97:** 309-313.