Journal of Sciences, Islamic Republic of Iran 36(1): 5 - 19 (2025) http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

Genetic Analysis of Y-STRs in Two Iranian Sub-Populations

M. Ghaderi-Zefrehei[1], F. Rafeie[2], A. Alipour Tabrizi[3], Z. Baratieh[4], S. Kamal Fotouhi[4], J. Smith[5], M. Muhaghegh Dolatabady[1*]

[1] *Department of Animal Science, Faculty of Agricultural Sciences, Yasouj University, Yasouj, Islamic Republic of Iran*
[2] *Department of Agricultural Biotechnology, Faculty of Agricultural Sciences, University of Guilan, Rasht, Islamic Republic of Iran*
[3] *Molecular Genetics Department, Mashhad Forensic Medical Center, Mashhad, Islamic Republic of Iran*
[4] *Forensic Medical Research Center, National Forensic Medical Organization, Tehran, Islamic Republic of Iran*
[5] *The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, UK*

Received: 3 December 2024 / Revised: 7 October 2025 / Accepted: 8 November 2025

Abstract

This study presents a comprehensive genetic analysis of 17 Y-chromosomal short tandem repeat (Y-STR) loci in two Iranian sub-populations from the Fars (n=109) and Isfahan (n=180) provinces. The loci investigated included DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635 (Y-GATA-C4), and Y-GATA-H4. Results demonstrated that the DYS385a/b locus exhibited the greatest allelic diversity in both populations, with 11 distinct alleles detected and mean allele counts of 6.29 and 5.88 in the Fars and Isfahan groups, respectively. Conversely, the Fars cohort showed the lowest allelic variation (three alleles) at DYS439 and Y-GATA-H4 loci, while the Isfahan population exhibited minimal variation (four alleles) at DYS19 and DYS439. Haplotype analyses revealed intra-population sharing rates of 2.75% in Fars and 10.0% in Isfahan, with an overall 8.3% haplotype overlap observed across the combined dataset of 289 individuals. Both populations exhibited high haplotype diversity values approaching 0.99, indicating substantial genetic variability. The haplotype discrimination capacity varied among populations, with value of 0.9725 for Fars, 0.8519 for Isfahan, and 0.9170 for the entire sample set. Population differentiation was assessed using pairwise FST and RST metrics, which confirmed significant genetic divergence between Fars and Isfahan groups (FST = 0.00743, $p < 0.001$; RST = 0.0106, $p < 0.01$). These findings underscore the genetic distinctness of the two sub-populations. The study highlights the necessity for further research incorporating Y-chromosomal single-nucleotide polymorphisms (Y-SNPs), larger sample sizes, and additional ancestral information to enhance the understanding of genetic structure and demographic history within Iranian populations.

Keywords: Haplotype; Genetic diversity; Y-STR markers.

* Corresponding Author: Tel: +98 74331006000; Email: mmuhaghegh@yu.ac.ir

Introduction

The genome of most organisms contains repetitive DNA sequences that are very small in length (2 to 7 bp), known as microsatellites, or commonly called short tandem repeats (STRs) (1). The STRs have become popular among repetitive DNA markers because of high variability in their repeat numbers and easy determination of their genotypes (2). The STR loci, as a genetic marker, have been widely used for individual identification (3, 4), population characterization (5), molecular diagnosis of genetic diseases (6), and even forensic studies (7). The STRs on the Y chromosome and its associated Y-STR haplotype, makes it potentially valuable for forensic genetic particularly in sexual assault investigations (8). Males commit the vast majority of violent crimes; therefore, a crime-scene sample left by a culprit will usually be informative when typed with Y chromosome-specific markers (9), and even if no Y-specific markers were found in a sample, this result would still be helpful. In addition, Y-STR analysis readily provides an assailant-specific profile in cases of Azoospermia (10).

Iran is a heterogeneous country that makes up about 1.12% of the total world population (https://www.worldometers.info/) has significant diversity due to factors such as climate, agriculture, industry, ethnicity, religion and language (11). These factors lead to significant genetic diversity, although the Iranian population is studied poorly, particularly concerning Y-STR data. Mixed signals of gene diversity are observed in Iranian populations. For example, no significant difference has been observed in genetic distance between the Iranian Gilak, Mazandarani, Bakhtiari, and non-Iranian Turkish, Azerbaijanian, Armenian, and Kurdish ethnic groups using Y-STRs. However, a significant gene distance was observed among Talysh, the Iranian Arabs, Georgian and Kazakh ethnic groups (12). A total of 204 unique Y-STR haplotypes were detected in unrelated male samples from Mazandaran and Gilan provinces. However, no unique Y-STR haplotypes were observed across the two provinces (13). The results of a study showed the efficiency of Y-STRs as a reliable tool in forensic genetic medicine and the evolutionary comparison of different ethnic groups from northern Iran (14). In addition, in another study, based on whole genome data from eleven ethnic groups present in Iran, it has been observed that the Iranian population consists of genetically clusters of overlapping ethnic groups, several strongly admixed groups, and shows distinct genetic diversity with respect to populations in close geographic proximity (15). In the current study, the Y-STR profiles were investigated in two Iranian sub-populations (Isfahan and Fars provinces). These two regions cover a large section of the ancient capitals of the Persian Empire, in which population genetics profile analysis has rarely been undertaken. The study had two primary aims. First, forensics, to provide a database with high-quality haplotypes with more reliable Y-STRs markers than currently available studies which have been conducted in these regions. Second, anthropological, as the data and results can potentially be reused by any human population anthropologist to reconstruct the history of the Iranian or surrounding populations.

Materials and Methods

Populations and DNA profiling

Through confirmation of ethnic identity (due to the mixed nature of the two populations via continual migration), only healthy and unrelated male participants capable of tracing their ancestry for at least two generations were included. The target sample sizes of n = 109 for Fars and n = 180 for Isfahan were determined a priori using G*Power 3.1 software to ensure ≥80% statistical power ($\alpha = 0.05$) to detect a medium effect size ($f = 0.25$) in haplotype diversity differences between the two populations, based on previously reported genetic variation in Iranian subpopulations (13). This calculation ensured sufficient sensitivity to detect meaningful population differentiation while accounting for the expected within-population heterogeneity inherent in Y-STR markers. Y-DNA samples from two socio-geographical populations in the Fars (109 samples) and Isfahan (180 samples) provinces were collected by the Iranian Medical Research Council and the Iranian Forensic Genetics Bank, in collaboration with two referral centers in the capital cities of the two provinces. DNA was isolated from bloodstains on 1.2 mm FTA cards using the QIAamp DNA Investigation Kit (Qiagen). Multiplex polymerase chain reaction (PCR) was performed to amplify the 17 Y-STR loci (DYS19, DYS385a, DYS385b, DYS389I, DYS389II, DYS390,

DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635/Y-GATA-C4, and Y-GATA-H4) using the commercially available AmpFLSTR® Yfiler™ PCR Amplification Kit (Applied Biosystems, Foster City, CA). This kit employs a pre-designed, multiplexed primer set optimized for human Y-chromosome STR analysis; the primer sequences are proprietary and not publicly disclosed by the manufacturer. However, the amplicon sizes, fluorescent labeling (6-FAM, NED, PET, VIC), and amplification conditions are standardized and validated according to the manufacturer's specifications (Applied Biosystems). All reactions were carried out under the following thermal cycling parameters: initial denaturation at 95°C for 11 min, followed by 28 cycles of 94°C for 1 min, 58°C for 1 min, and 72°C for 1 min, with a final extension at 60°C for 45 min. Multiplex PCR products were analyzed on an automated ABI 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA) using POP-4 polymer in a 36 cm capillary array. Quality control for Y-STR typing was confirmed by successful genotyping of five blind reference samples provided by the Y-Chromosome Haplotype Reference Database (YHRD; https://yhrd.org/) (16). Haplotypes were interpreted using GeneMapper ID-X Software v1.3 (Applied Biosystems), and allele calling was based on the Nei equation (17). The obtained Y-STR data were submitted to the YHRD under Accession Numbers YA004229 (Fars) and YA004228 (Isfahan). The general formal ethical approval for this study was obtained from the Ethical Committee of the Legal Medicine Organization, Tehran (Reference ID: 1396.0018).

Y-STR data analyses

GenAlEx V6.5 (18) was employed to determine allelic frequencies, number of alleles, number of matches between haplotypes, and Shannon index for the 17 Y-STR markers. The Shannon's Index is an information statistic index, assuming all types are represented in a sample and that they are randomly sampled. It combines both evenness and richness in a single measure (19). Gene diversity was also calculated based on the Nei equation (17). Also, shared haplotypes, theta parameters (θ; the number of mutations per generation, assuming that the population is in mutation-drift equilibrium), and expected heterozygosity (HE) were estimated using Arlequin software (20). Theta was calculated from the expected homozygosity (HomE) as follows:

$$\theta = \frac{1\text{-HomE}}{\text{HomE}}$$

where, HomE = 1 − HE. Moreover, haplotype discrimination capacity (DC) was calculated as follows (21):

$$DC = \frac{\text{Number of unique haplotypes}}{\text{Number of total genotype markers}}$$

Additionally, the probability of exclusion (PE) was estimated by (22):

$$PE = 1\text{-} \sum P_i^2$$

where P is the haplotype frequency.

HapYDive software (http://www.portugene.com/hapydive.html) (23) was also used to calculate the number of different haplotypes and haplotype diversity (HD). The haplotype match probability (HMP) was calculated as $HMP = 1 - HD$. Haplogroup Predictor software (www.hprg.com/hapest5/hapest5a/hapest5.htm) was used to predict the haplogroup from the haplotypes (24,25). The 17-marker haplotypes were analyzed using the following 26 haplogroups (as defined by ISOGG 2017 Y Tree): C2-M217, D-M174, E1a-M132, E1b1a-V38, E1b1b-M35, G1-M342, G2a-P15, G2b-M377, H-M69, I1-M253, I2a-P37(xM26), I2a-M26, I2b-M223(xM284), I2b-M284, J1-M267, J2a-M410, J2b-M12, L-M11, N-M46, O1-F265, O2-M122, Q-M242, R1a-M17, R1b-M349, R2-M124, and T-M184. Generally, a probability of 85% with a fitness score of 20 was considered acceptable for prediction. In a few cases, a fractional result was obtained for STRs, but since the Haplogroup Predictor software cannot handle such values, the values were truncated before being analyzed. Analysis of Molecular Variance (AMOVA) and of F_{st} and R_{st} values were calculated using GenAlEx V6.5 and YHRD (Release 59- 2018/Nov/01) softwares (p-values computed by 10,000 permutations) for obtaining the 17 Y-STR haplotypes and also data from neighboring countries, including Pakistan (Pathan), Turkey, and Iraq (Table S1).

Results

Despite the existence of manuscripts with larger sample sizes exploring Iranian subpopulations, this paper stands out for explicitly focusing on two key Iranian capital regions that were once capitals of the Old Persian Empire. These regions cover a significant portion of the Iranian plateau. Furthermore, we haven't only conducted a forensic genetic analysis but also delved into demographic analysis using Y-STR data.

In the current study, the Y-STR profiles were analyzed in two Iranian subpopulations (Isfahan and Fars provinces), including 289 random, unrelated forensic samples. Haploid allelic frequencies and number of alleles for the 17 Y-STR loci in Fars and Isfahan populations are shown in Figure 1 and Table S2. The results of estimated genetic diversity parameters are shown in Figure 2a, b, c and d and Table S3. The

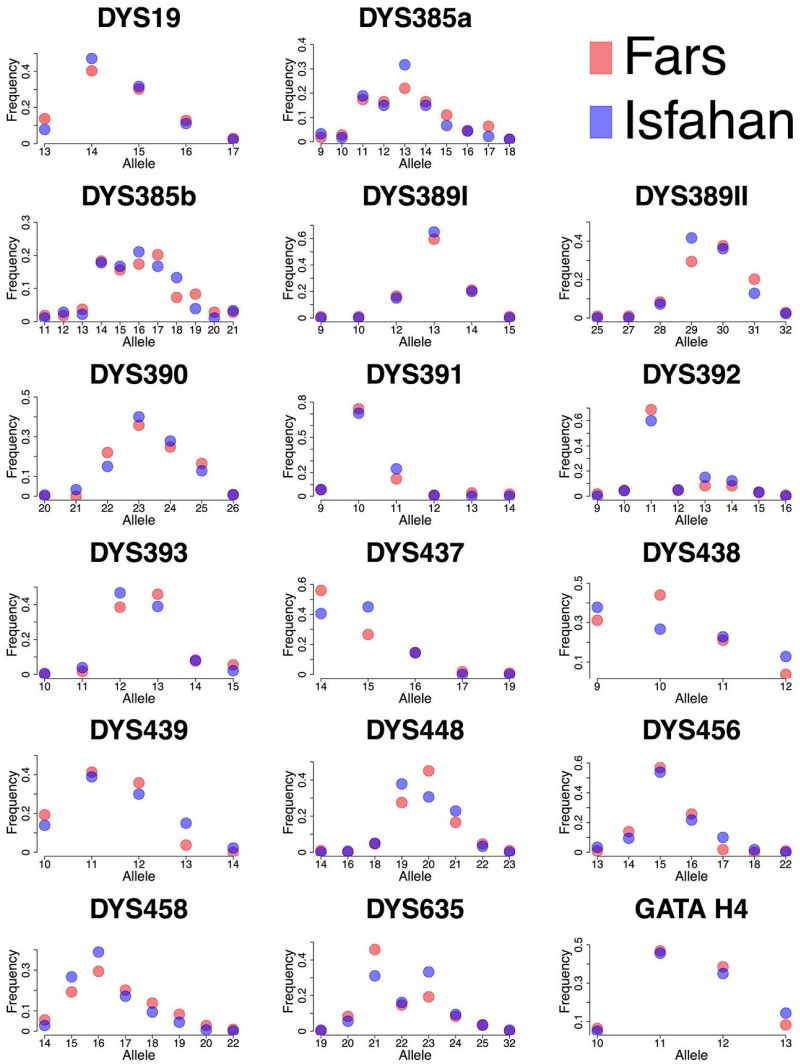

Figure 1. Haploid allelic frequencies of 17 STRs in Fars and Isfahan populations are presented in red and blue, respectively (purple represents overlap). For detailed information about each population, see Tables S1 and S2.

maximum number of alleles (11 alleles) was found in the DYS385a/b locus in both populations. The minimum number of alleles was observed for the DYS439 and GATA-H4 loci, 3 and 4 alleles each in Fars and Isfahan, respectively (Figure 2a). For further analysis of the DYS385a/b locus, the Shannon index values for Fars (2.092) and Isfahan (2.025) populations and the HE values (0.86 for Fars and 0.85 for Isfahan) were calculated (Figure 2b and c, and Table S3). In samples from the Fars province, the highest values for the HE parameter were detected in DYS385b (0.86), DYS385a (0.85), DYS458 (0.81), DYS390 (0.74), and DYS389II (0.73). Similarly, these values for Isfahan samples were in DYS385b (0.85), DYS385a (0.81), DYS635 (0.75), DYS390 (0.74), and DYS458 (0.72). The highest values of Theta (θ) for DYS385b were 26 in Fars and 22 in Isfahan individuals, followed by 23 (DYS385a) and 13 (DYS458) in Fars and 14 (DYS385a) and 7 (DYS635) in Isfahan samples (Figure. 2d). Pairwise population differences were based on the distance method and were used to calculate F_{st} values. Even though a comparative analysis of the Isfahan population indicated that the population is slightly different from the Fars population (Fst = 0.00743, $p<0.001$), it is hardly can be said it is statistically different as its values obtained are low and less than that 0.15. The R_{st} value obtained for the Isfahan

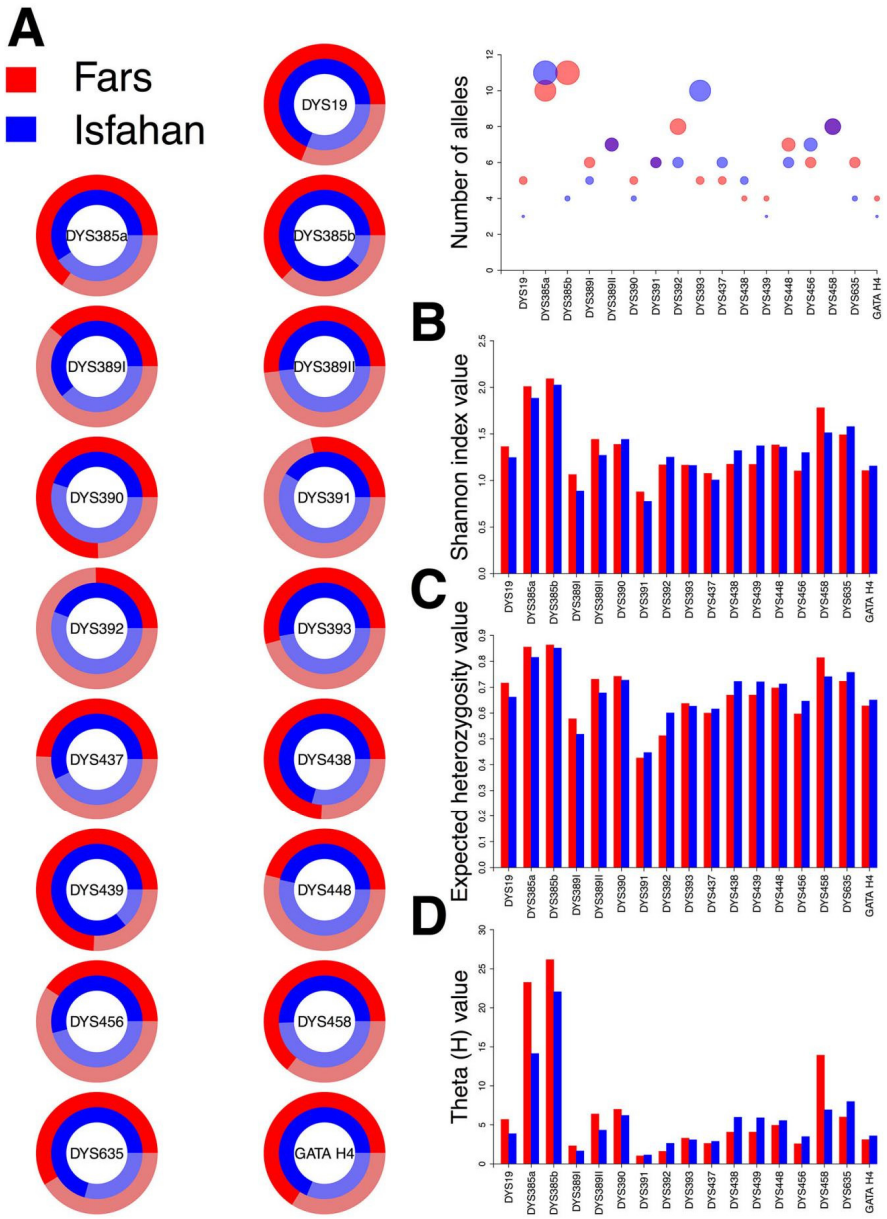

Figure 2. Number of alleles, gene diversity and theta values of 17 STRs in Fars (red) and Isfahan (blue) populations, Iran (purple represents overlaps). For more information about each population, see Table S3. A: Daunt charts show total number of alleles in each STR where the effective number of alleles is shown in bold and non-effective are shown in pale colours. Circle size of the corresponding plot represents number of alleles for each STR of the studied populations. DYS389II and DYS391 had the exact same number of alleles in both Fars and Isfahan populations (7 and 6, respectively). B: Shanon index, C: Expected heterozygosity and D: Theta (θ).

and Fars provinces was 0.0.0106 (p<0.01). The numbers of total and unique haplotypes in each studied population, along with the number of haplotypes shared within and between the Fars and Isfahan populations, are shown in Table 1. The shared haplotypes were either cases of two or three matching haplotypes. There were 106 distinct haplotypes within the Fars population, while three of the haplotypes occurred twice. There were 161 distinct haplotypes within the Isfahan population while 3 and 16 haplotypes occurred three and twice times, respectively.

Table 1. Haplotype and forensics parameters for 17 Y-STR loci deriving from 289 samples from Isfahan and Fars provinces of Iran

Parameters	Fars	Isfahan	Total Population
Number of haplotypes	109	189	289
Number of Different haplotypes	106	161	265*
Shared haplotypes			
Cases of 3 matching haplotypes	0	3 (1.58%)	3 (1.03%)
Cases of 2 matching haplotypes	3 (2.75%)	16 (8.46 %)	19 (6.57%)
Total shared haplotypes	3 (2.75%)	19 (10.05%)	24 (8.30%)
Forensic parameters			
Haplotype discrimination capacity (DC)	0.9725	0.8519	0.9170
Chance of exclusion (CE)	0.6676	0.6720	0.6698
Haplotype diversity (HD)	0.9995	0.9984	0.9989
Gene diversity	0.6737	0.6758	0.6774
Haplotype match probability (HMP)	0.0005	0.0016	0.0011

* Two haplotypes were shared between Fars and Isfahan populations, so the number in total was reduced by 2

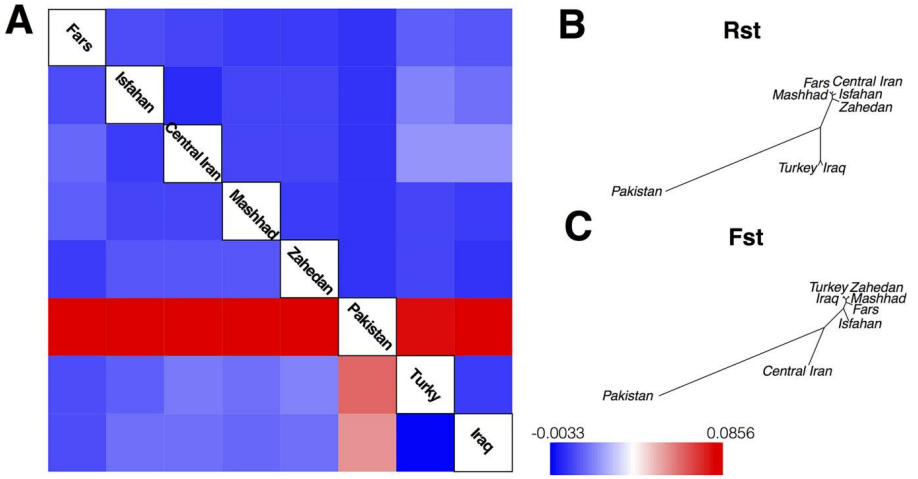

Figure 3. The AMOVA analysis (F_{st} and R_{st} values) of the Y-Chromosome Haplotype Reference (YHRD (https://yhrd.org/) (3); *P*-values computed by 10,000 permutations). The off-diagonal entries in the matrix provide the F_{st} and R_{st} values between each variable pair. For more information about each population see Table S1. A: The values below the diagonal are R_{st} and values above the diagonal are F_{st} obtained for comparison between five Iranian and three regional populations (Tables S4 and S5), B: The unrooted tree of R_{st} values of five Iranian and three regional populations (Table S4) C: The unrooted tree of F_{st} values obtained for comparison between five Iranian and three regional populations (Table S5).

Two haplotypes from Fars matched with two haplotypes from Isfahan, but did not match with any other haplotypes within the populations. In total, 24 (8.30%) of the 289 haplotypes were shared between and within subpopulations. The population genetics parameters DC, CE, HD, gene diversity, and HMP were also analyzed (Table 1). Different populations from Iran and neighboring countries were used for regional analysis (Table S1). The highest and lowest R_{st} values were seen for Pakistan-Fars (0.0856) and Turkey-Iraq (-0.0008), respectively (Figure 3a and b and Table S4). The highest and lowest F_{st} values were observed for Central Iran-Isfahan (-0.0033) and Pakistan-Central Iran (0.0056), respectively, as shown in Figure 3a and c and Table S5. Allelic frequencies can be different between different Y haplogroups. Indeed, those differences form the basis for haplogroup prediction from Y-STR values. Table 2 shows the predicted haplogroup frequencies for the Fars and Isfahan samples, plus the summary for the combined populations. Note that the SNP names in Table 2 are only for haplogroup identification purposes - no SNPs were tested in the current study. As can be seen in Table 2, the highest haplogroup frequency was observed for J2a and R1a haplogroups. Unfortunately, in 12 out of 280 haplotypes, no haplogroup prediction could be made with the set of haplogroups employed in the prediction

Table 2. Haplogroup Summary Statistics

All Samples			Fars Samples			Isfahan Samples		
Haplogroup	Count	Percent	Haplo	Count	Percent	Haplo	Count	Percent
C2-M217	2	0.7%	C2-M217	2	1.8%	C2-M217	0	0.0%
D-M174	1	0.3%	D-M174	1	0.9%	D-M174	0	0.0%
E1a-M132	0	0.0%	E1a-M132	0	0.0%	E1a-M132	0	0.0%
E1b1a-V38	4	1.4%	E1b1a-V38	0	0.0%	E1b1a-V38	4	2.2%
E1b1b-M35	23	8.0%	E1b1b-M35	13	11.9%	E1b1b-M35	10	5.6%
G1-M342	6	2.1%	G1-M342	2	1.8%	G1-M342	4	2.2%
G2a-P15	15	5.2%	G2a-P15	9	8.3%	G2a-P15	6	3.3%
G2b-M377	0	0.0%	G2b-M377	0	0.0%	G2b-M377	0	0.0%
H-M69	10	3.5%	H-M69	5	4.6%	H-M69	5	2.8%
I1-M253	1	0.3%	I1-M253	1	0.9%	I1-M253	0	0.0%
I2a-P37(xM26)	3	1.0%	I2a-P37	2	1.8%	I2a-P37	1	0.6%
I2a-M26	0	0.0%	I2a-M26	0	0.0%	I2a-M26	0	0.0%
I2b-M223(xM284)	0	0.0%	I2b-M223	0	0.0%	I2b-M223	0	0.0%
I2b-M284	1	0.3%	I2b-M284	0	0.0%	I2b-M284	1	0.6%
J1-M267	15	5.2%	J1-M267	10	9.2%	J1-M267	5	2.8%
J2a-M410	74	25.6%	J2a-M410	20	18.3%	J2a-M410	54	30.0%
J2b-M12	5	1.7%	J2b-M12	0	0.0%	J2b-M12	5	2.8%
L-M11	19	6.6%	L-M11	4	3.7%	L-M11	15	8.3%
N-M46	3	1.0%	N-M46	2	1.8%	N-M46	1	0.6%
O1-F265	1	0.3%	O1-F265	0	0.0%	O1-F265	1	0.6%
O2-M122	2	0.7%	O2-M122	1	0.9%	O2-M122	1	0.6%
Q-M242	11	3.8%	Q-M242	5	4.6%	Q-M242	6	3.3%
R1a-M17	37	12.8%	R1a-M17	16	14.7%	R1a-M17	21	11.7%
R1b-M343	22	7.6%	R1b-M343	3	2.8%	R1b-M343	19	10.6%
R2-M124	12	4.2%	R2-M124	3	2.8%	R2-M124	9	5.0%
T-M184	10	3.5%	T-M184	6	5.5%	T-M184	4	2.2%
Unknown	12	4.3%	Unknown	4	3.7%	Unknown	8	4.4%
Total	289		Total	109		Total	180	

program. This is probably due to four main factors: 1) the relatively small number of Y-STR markers (17 Y-STRs), 2) the omission in the set of 17 Y-STRs of markers of particular importance to haplogroup prediction (DYS388, YCAII), 3) the likelihood that the haplotypes in these studied populations have diverged somewhat from the populations represented in the predictor program database, and 4) the possible existence in these populations of haplogroups that are not included in the analysis. A total of 22 haplogroups were found in the two populations. About a quarter of the total population was found to be in the J2a-M410 haplogroup, while the next most common haplogroup was Haplogroup R1a-M17 at about 13%. In addition, 7.3% of the combined populations were found to be in haplogroup R1b-M343, 7.3% in E1b1b-M35, 7.3% in J1-M267, 6.7% in L-M11, 4.7% in G2a-P15, 4.2% in R2-M124, 3.9% in Q-M242, and 2.8% in H1-M69. The remaining haplogroups occurred at frequencies of 2% or less.

Discussion

Different ethnic groups, religious affiliations, languages, climates, *etc.,* have produced a heterogeneous population in Iran. Thus, exploring Y-STRs in Iran is a compelling venture. Both F_{st} (the classical allele frequency-based differentiation estimator) and R_{st} (the mutational based model differentiation estimator) measures, though significant, showed low values in this study, indicating that genetic variance of within-population is much larger than between population. Many Y-STRs-based studies have reported both F_{st} and R_{st}. For example, the most significant genetic distance (R_{st}=0.35434) was observed in Gilaks and Azeris; but, Sistanis and Lurs groups showed the smallest genetic distance (R_{st}=0.00483) compared to other Iranian ethnicities (26). However, there is a longstanding debate about using both measures in terms of their variance sampling, mutational model, distance-related detection, expectation of new genetic mutation, etc. In other words, no mutation model completely matches the behavior of microsatellites and no clear consensus over the accuracy of these measures to estimate genetic distances between populations, therefore both F_{st} and R_{st} measures, are often used together in microsatellite data analyses (27). Since both measures have their own drawbacks, we wanted these measures to complement each other, and so both methods were used in this study.

Genetic diversity parameters distinguished that DYS385a/b and DYS439 loci revealed the maximum and minimum number of alleles in both populations, respectively. Due to the lack of uniformity of the marker set used in different studies, the results are not very comparable. DYS385 was reported to show the most significant number of alleles in the Finnish population using 16 Y-STR loci (22). Also, the highest and lowest number of alleles were found for DYS458 and GATA-H4 markers, respectively, in ethnic Malaysian groups (Malay, Hindi, and Chinese) (28). Additionally, using a set of 11 Y-STR loci with demographic groups in America, the maximum number of alleles (18 alleles) was found in DYS385a/b markers, and the lowest number of alleles (5 alleles) was observed in DYS389I (29). Considering the DYS385a/b locus, the Shannon index values and the HE values were also calculated. The highest HE and Theta (θ) values were found for DYS385b and DYS385a loci in both populations. The comparative analysis, via F_{st} pairwise distance, revealed the statistical difference between the Fars and Isfahan populations. More than 8% of the 289 haplotypes were shared between and within the sub-populations. The American ethnic groups (Spanish, African, and European) were studied using 7 Y-STRs and determined that 85, 83, and 82% of the haplotypes were respectively observed only once in each population (30). The male population of Turkey was studied using 9 Y-STRs and most haplotypes were observed only once in the population (31). The average value of CE (0.6698), high value of HD, maximum (0.9725 in Fars), and minimum (0.8519 in Isfahan) DC values were compared with other corresponding studies. In another study, using 9 Y-STRs, the obtained HD and DC values were as 0.9938 and 8, respectively, in the Isfahan population, which is somewhat higher than the current results (32). Reports from other parts of Iran show that Tehran had HD and DC values of 0.997 and 0.900, respectively (32), while the HD and DC values were 0.9999 and 0.9884, respectively, in the Eastern region of Iran (33). Investigation of 17 Y-STRs in four Pakistan ethnic groups (Balochi, Pathans, Punjabis, and Sindhi) revealed the highest HD (HD=0.9957) for a Pathan population with 102 haplotypes (34). Using 27 Y-STRs, 102 haplotypes were identified in Kazakh populations and the HD and DC were 0.997 and 0.91, respectively (35). Furthermore, a total of 224 haplotypes were detected in two Mestizo populations, of which 98.66% were unique (36).

In addition, socio-geographical effects have been reported based on Y-STR diversity in different ethnic groups in America, including African, European, Spanish, Native, and Asian ethnic groups from various geographical regions. The highest DC was found in America's Asian population from two metropolitan areas, Tucson, AZ, and New York, NY. The Asian population in America also showed the most significant amount of HD (0.84) compared to the other ethnic groups (37). The average values of DC and HD were 0.9060 and 0.9994, respectively, in the European part of Russia (38). In a study on Indian population groups using 5 Y-SNPs markers, HD was in the range of 0.70 to 1.00 (39), while in another study, the average HD was 0.98 (40). To create a Y-STR database for the American population, the estimated DC values were 0.74, 0.75, and 0.69 for Hispanic, African, and native American populations, respectively (41). Using Y-SNPs and Y-STRs analysis, the HD and DC values were 0.9998 and 0.9680, respectively, in the Hui population of Liaoning province (42).

In terms of population genetic structure, our Y-STR data revealed minor differences between the two population groups, with the Isfahan population exhibiting lower levels of within-population diversity, higher levels of among-population differentiation, and a tendency toward a lower rate of demographic growth compared to the Fars population (Table S3). Comparing different Iranian populations with neighboring countries showed that the south Iranian populations have been influenced by eastern and north-western neighbors, with Pakistan having a larger effect on the south Iranian gene pool than Turkey (43). In a fascinating study, the northwest region of Iran was investigated using haplogroup G2-M406 (44). This region is renowned as the cradle of great civilizations like Mesopotamia and Elam, which is surrounded by the Zagros Mountains. It was calculated that the emergence of haplogroup G2-M406 in this region dates back more than 8,800 years. They highlighted that the gene flow from neighboring areas and the free movements of the local people had been sharply restricted by geography due to the Zagros Mountains. They also showed a westward migration of the Iranian population dwelling in this plateau to Europe through Turkey and the Caucasus (44). Haplogroup G has arisen due to a mutation from Haplogroup F, M201 in the case of Haplogroup G, and it has been reported that it has low frequencies in Western Europe (45). In haplogroup G, the segregation of G1 and G2 subdivisions occurred about 19,000 years ago, and the G2a and G2a3b subdivisions about 15,000 and 12,600 years ago, respectively. Although G2a was formed about 15,000 years ago, its contemporary subdivision expanded about 5,600 years ago from the Middle East into two branches: one to the Balkans and Western Europe, the other to the Caucasus and Pakistan (46). Various studies have addressed the origin of haplogroup G being in the regions between

eastern Anatolia, Armenia, and western Iran, with most of its current expansion being in the Middle East, the Caucasus, and southern Europe (Italy and Greece) (47).

Even having an average mutation rate of 0.2% per generation, Y-STRs can demonstrat the patrilineal affinities among subpopulations, and reveal the genetic composition and consanguinity scenarios (48), and are widely used to predict haplogroups with online tools. However, the occurrence of errors in Y-STR haplogroup prediction should not be overlooked. To this end, by increasing the number of Y-STRs employed to predict the haplogroup, correcting inaccurate prediction due to mutation is necessary using precise Y-SNP data. In other words, the existence of errors in Y-STR haplogroup predictions implies that validation using SNP analysis is appropriate when high accuracy is required in terms of haplogroup prediction (49). It has been shown that applying Y-SNPs in combination with Y-STR data can help shed light on the biogeographical origins of unknown male individuals with high precision, as this could mediate accessibility to robust phylogenetic trees and large-scale genome reference data (50). Therefore, more studies are needed to provide more insightful information about the geographic distribution of many Y-SNPs, and these sets of data could improve the geographic resolution of paternal ancestry inferences. One of the prime limitations of this study was the magnitude of subpopulation admixture and inter-ethnic structure. This matter wasn't considered in this study. Both inter-ethnic admixture and ancestral source subpopulations may contribute to fine scale Y-STRs heterogeneity within Iran ethnic groups. For the future genetic investigation of Iran's ethnic groups and subpopulation admixture, a panel of Y-SNPs should be considered, as it could infer distant evolutionary ancestry in comparison to Y-STRs that reflect close familial kinships.

Conclusion

This study covers comparative statistical analysis and various population genetics parameters from two socio-geographical populations of Iran (Fars and Isfahan) using 289 random unrelated samples and 17 Y-STRs as markers. The DYS385a/b marker revealed the maximum number of alleles in both studied populations. However, DYS439 and GATA-H4 markers in the Fars population and DYS19 and DYS439 markers in Isfahan individuals showed the minimum number of alleles. A high rate of haplotype diversity was observed in the total population and subpopulations of the Fars, and Isfahan regions. Haplotype discrimination capacity varied in Fars, Isfahan and the total populations. The results, again, showed the power of Y-STR markers for forensic national/regional human identification and haplotype comparison purposes.

Acknowledgements

We would like to express our gratitude to all the contributors, sample donors, and our colleagues in the forensic centers of the capital cities of Fars and Isfahan provinces. We are deeply indebted to Amin Mortazavi, Reza Raoofian, Reza Pazhoomand, Whit Athey, Farzan, Ghane Golmohammadi for their supports and help in this research.

Ethical Approval

The authors are committed to all ethical standards for human studies. This research protocol was under supervision by the Ethical Review Committees of Iranian Legal Medicine Research Center. The general formal ethical approval for this study was obtained from the Ethical Committee of the Legal Medicine Organization, Tehran (Approval ID: IR.LMO.REC.115021). Informed consent was obtained from all individual participants included in the study for using samples and publishing results.

References

1. Ellegren H. Microsatellites: simple sequences with complex evolution. Nature Reviews Genetics. 2004;1;5(6):435-45.
2. Butler JM, Hill CR. Biology and genetics of new autosomal STR loci useful for forensic DNA analysis Forensic Science Review. 2012;24(1):15.
3. McClintock JT. Using conventional STR technology in analyzing biological transfer evidence. InNext Generation Sequencing (NGS) Technology in DNA Analysis 2024 ;1 (pp. 21-34). Academic Press.
4. Vallian S, Moeini H. Genotyping of Five Polymorphic STR Loci in Iranian Province of Isfahan. Journal of Sciences, Islamic Republic of Iran. 2006;17(2):113-117
5. Agudo MM, Aanes H, Albert M, Janssen K, Gill P, Bleka Ø. An overview of autosomal STRs and identity SNPs in a Norwegian population using massively parallel sequencing. Forensic Science International: Genetics. 2024; 3:103057.
6. Soloveva EV, Skleimova MM, Minaycheva LI, Garaeva AF, Zhigalina DI, Churkin EO, Okkel YV, Timofeeva OS, Petrov IA, Seitova GN, Lebedev IN. PGT-M for spinocerebellar ataxia type 1: development of a STR panel and a report of two clinical cases. Journal of Assisted Reproduction and Genetics. 2024;41(5):1273-83.
7. Pedroza Matute S, Iyavoo S. Applications and Performance of Precision ID GlobalFiler NGS STR, Identity, and Ancestry Panels in Forensic Genetics. Genes. 2024;28;15(9):1133.
8. Costa R, Fadoni J, Amorim A, Cainé L. Y-STR Databases—Application in Sexual Crimes. Genes. 2025;

25:16(5):484.
9. Jobling MA, Pandya AR, Tyler-Smith CH. The Y chromosome in forensic analysis and paternity testing. International Journal of Legal Medicine. 1997; 110:118-24.
10. Jobling MA, Tyler-Smith C. The human Y chromosome: an evolutionary marker comes of age. Nature Reviews Genetics. 2003;1;4(8):598-612.
11. Ahmady K. Ethnic, national and identity demand in Iran on the axis of justice and development: A grounded theory method study among five major ethnic groups. European Journal of Humanities and Social Sciences. 2023;31;3(5):41-54.
12. Roewer L, Willuweit S, Stoneking M, Nasidze I. A Y-STR database of Iranian and Azerbaijanian minority populations. Forensic Science International: Genetics. 2009; 1;4(1):e53-5.
13. Sayyari M, Salehzadeh A, Tabatabaiefar MA, Abbasi A. Profiling of 17 Y-STR loci in Mazandaran and Gilan provinces of Iran. Turkish Journal of Medical Sciences. 2019;49(5):1277-86.
14. Sayyari M, Salehzadeh A, Tabatabaiefar MA, Ali A. Genetic polymorphisms of Y-chromosome short tandem repeats (Y-STRs) in a male population from Golestan province, Iran. Molecular Biology Research Communications. 2020;9(1):11.
15. Mehrjoo Z, Fattahi Z, Beheshtian M, Mohseni M, Poustchi H, Ardalani F, et al. Distinct genetic variation and heterogeneity of the Iranian population. PLoS Genetics. 2019;24;15(9): e1008385.
16. Willuweit S, Roewer L. The new Y chromosome haplotype reference database. Forensic Science International: Genetics. 2015;1; 15:43-8.
17. Nei M. Molecular evolutionary genetics. Columbia university press; 1987;31.
18. Peakall RO, Smouse PE. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes. 2006;6(1):288-95.
19. Moges AD, Admassu B, Belew D, Yesuf M, Njuguna J, Kyalo M, et al. Development of microsatellite markers and analysis of genetic diversity and population structure of Colletotrichum gloeosporioides from Ethiopia. PloS One. 2016;15;11(3): e0151257.
20. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 1992;1;131(2):479-91.
21. Robino C, Crobu FR, Di Gaetano C, Bekada A, Benhamamouch S, Cerutti N, et al. Analysis of Y-chromosomal SNP haplogroups and STR haplotypes in an Algerian population sample. International Journal of Legal Medicine. 2008; 122:251-5.
22. Hedman M, Pimenoff V, Lukka M, Sistonen P, Sajantila A. Analysis of 16 Y STR loci in the Finnish population reveals a local reduction in the diversity of male lineages. Forensic Science International. 2004;28;142(1):37-43.
23. Alves C, Gusmão L, Meirinhos J, Amorim A. Making the most of Y-STR haplotypes: The HapYDive. In International Congress Series 2006 Apr 1 (Vol. 1288, pp. 201-203). Elsevier.
24. Athey TW. Haplogroup prediction from Y-STR values using an allele-frequency approach. Journal of Genetics and Genomics. 2005; 1:1-7.
25. Athey TW. Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach. Journal of Genetics and Genomics. 2006;2(2):34-9.
26. Oh A, Oh BU. Genetic differentiation that is exceptionally high and unexpectedly sensitive to geographic distance in the absence of gene flow: Insights from the genus Eranthis in East Asian regions. Ecology and Evolution. 2022;12(6): e9007.
27. Alinaghi S, Mohseni M, Fattahi Z, Beheshtian M, Ghodratpour F, Ashrafi FZ, Arzhangi S, Jalalvand K, Najafipour R, Khorshid HR, Kahrizi K. Genetic Analysis of 27 Y-STR Haplotypes in 11 Iranian Ethnic Groups. Archives of Iranian Medicine. 2024 Feb;27(2):79.
28. Chang YM, Perumal R, Keat PY, Kuehn DL. Haplotype diversity of 16 Y-chromosomal STRs in three main ethnic populations (Malays, Chinese and Indians) in Malaysia. Forensic Science International. 2007;22;167(1):70-6.
29. Shewale JG, Bhushan A, Nasir H, Schneida E, Washington B, Fleming A, et al. Population Data for Four Population Groups from the United States for the Eleven Y-Chromosome STR Loci Recommended by SWGDAM. Journal of Forensic Sciences. 2006;1;51(3).
30. Kayser M, Krawczak M, Excoffier L, Dieltjes P, Corach D, Pascali V, et al. An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. The American Journal of Human Genetics. 2001;1;68(4):990-1018.
31. Henke J, Henke L, Chatthopadhyay P, Kayser M, Dulmer M, Cleef S, et al. Application of Y-chromosomal STR haplotypes to forensic genetics. Croatian Medical Journal. 2001;1;42(3):292-7.
32. Ikram MS, Mehmood T, Rakha A, Akhtar S, Khan MI, Al-Qahtani WS, Safhi FA, Hadi S, Wang CC, Adnan A. Genetic diversity and forensic application of Y-filer STRs in four major ethnic groups of Pakistan. BMC Genomics. 2022;30;23(1):788.
33. Nasidze I, Schädlich H, Stoneking M. Haplotypes from the Caucasus, Turkey and Iran for nine Y-STR loci. Forensic science international. 2003;14;137(1):85-93.
34. Tabrizi AA, Hedjazi A, Kerachian MA, Honarvar Z, Dadgarmoghaddam M, Raoofian R. Genetic profile of 17 Y-chromosome STR haplotypes in East of Iran. Forensic Science International: Genetics. 2015;1;14: e6-7.
35. Ashirbekov Y, Zhunussova A, Abaildayev A, Bukayeva A, Sabitov Z, Zhabagin M. Genetic polymorphism of 27 Y-STR loci in Kazakh populations from Central Kazakhstan. Annals of Human Biology. 2024; 31:51(1):2377571.
36. Sánchez-Méndez AD, Narvaez-Rivera SE, Rangel-Villalobos H, Hernández-Bello J, López-Quintero A, Moreno-Ortíz JM, Ramos-González B, Aguilar-Velázquez JA. Genetic Diversity and Forensic Parameters of 27 Y-STRs in Two Mestizo Populations from Western Mexico. Genes. 2025; 19:16(3):352.
37. Hammer MF, Chamberlain VF, Kearney VF, Stover D, Zhang G, Karafet T, et al. Population structure of Y chromosome SNP haplogroups in the United States and forensic implications for constructing Y chromosome STR databases. Forensic Science International. 2006;1;164(1):45-55.
38. Roewer L, Willuweit S, Krüger C, Nagy M, Rychkov S,

Morozowa I, et al. Analysis of Y chromosome STR haplotypes in the European part of Russia reveals high diversities but non-significant genetic distances between populations. International Journal of Legal Medicine. 2008; 122:219-23.
39. Cordaux R, Bentley G, Aunger R, Sirajuddin SM, Stoneking M. Y-STR haplotypes from eight south Indian groups based on five loci. Journal of Forensic Sciences. 2004;1;49(4):847-8.
40. Ayadi I, Ammar-Keskes L, Rebai A. Haplotypes for 13 Y-chromosomal STR loci in South Tunisian population (Sfax region). Forensic Science International. 2006;20;164(2-3):249-53.
41. Kayser M, Brauer S, Schädlich H, Prinz M, Batzer MA, Zimmerman PA, Boatin BA, Stoneking M. Y chromosome STR haplotypes and the genetic structure of US populations of African, European, and Hispanic ancestry. Genome Research. 2003;1;13(4):624-34.
42. Fu D, Adnan A, Yao J, Aldayan NH, Wang CC, Hongyi C. Unraveling the paternal genetic structure and forensic traits of the Hui population in Liaoning Province, China using Y-chromosome analysis. BMC Genomics. 2023;17;24(1):691.
43. Regueiro M, Cadenas AM, Gayden T, Underhill PA, Herrera RJ. Iran: tricontinental nexus for Y-chromosome driven migration. Human Heredity. 2006;16;61(3):132-43.
44. Bahmanimehr A, Nikmanesh F, Yepiskoposyan L. Paternal genetic landscape of Zagros region of Iran and its role in the gene flow in the populations of the region. Iranian Journal of Epidemiology. 2014;10(3):43-53.
45. Goff PG, Athey TW. Diagnostic Y-STR markers in haplogroup G. Journal of Genetic Genealogy. 2006;2(1):12-7.
46. Rootsi S, Myres NM, Lin AA, Järve M, King RJ, Kutuev I, et al. Distinguishing the co-ancestries of haplogroup G Y-chromosomes in the populations of Europe and the Caucasus. European Journal of Human Genetics. 2012;20(12):1275-82.
47. Balanovsky O, Zhabagin M, Agdzhoyan A, Chukhryaeva M, Zaporozhchenko V, Utevska O, et al. Deep phylogenetic analysis of haplogroup G1 provides estimates of SNP and STR mutation rates on the human Y-chromosome and reveals migrations of Iranic speakers. PLoS One. 2015;7;10(4):e0122968.
48. Tariq M, Ahmad H, Hemphill BE, Farooq U, Schurr TG. Contrasting maternal and paternal genetic histories among five ethnic groups from Khyber Pakhtunkhwa, Pakistan. Scientific Reports. 2022;19;12(1):1027.
49. Petrejčíková E, Hronská D, Bernasovská J, Boroňová I, Gabriková D, Bôžiková A, et al. Y-SNP analysis versus Y-haplogroup predictor in the Slovak population. Anthropologischer Anzeiger. 2014;1;71(3).
50. Roewer L. Y-chromosome short tandem repeats in forensics—Sexing, profiling, and matching male DNA. Wiley Interdisciplinary Reviews: Forensic Science. 2019;1(4): e1336.

Table S1. Each haplotype information used in current study

Region	Number of haplotypes	Accession number	Source
Fars[1]	109	YA004229	YHRD[2] – Current study
Isfahan[3]	180	YA004228	YHRD – Current study
Center of Iran	154	YA003782	YHRD
Mashhad[4]	127	YA003903	YHRD
Zahedan[5]	102	YA003901	YHRD
Pakistan[6]	269	YA003846	YHRD
Turkey[7]	296	YA002964	YHRD
Iraq[8]	124	YA003858	YHRD

[1] Fars province, one of the southern provinces of Iran, had 109 Y-DNA samples in current study. Fars province, the capital of ancient Iran, with 122608 Km^2 total area, and a population of 4851274 individuals (national statistic center of Iran in 2016 census, https://www.amar.org.ir/english) is the 4th most populated province.
[2] Y-Chromosome Haplotype Reference Database (YHRD - https://yhrd.org/)
[3] Isfahan province, one of the central provinces of Iran, had 180 Y-DNA samples in this study. Isfahan province with 107029 Km^2 total area, and a population of 5120850 individuals (national statistic center of Iran in 2016 census, https://www.amar.org.ir/english) is the 3rd populated province.
[4] Mashhad city is the capital city of Khorasan-e-Razavi province, one of the eastern provinces of Iran. Khorasan-e-Razavi province with 118884 Km^2 total area, and a population of 6434501 individuals (national statistic center of Iran in 2016 census, https://www.amar.org.ir/english) is the 2nd populated province.
[5] Zadehan city is the capital city of Sistan and Baluchestan province, one of the eastern provinces of Iran. Sistan and Baluchestan province with 181785 Km^2 total area, and a population of 27750414 individuals (national statistic center of Iran in 2016 census, https://www.amar.org.ir/english) is the 10th populated province.
[6] Pathan, Pakistan. Pakistan is one of the eastern neighbors of Iran with 959 km border line.
[7] Turkey is the north-western neighbor of Iran with 534 km border line.
[8] Iraq is the western neighbor of Iran with 1599 km border line.

Table S2. Haploid allelic frequencies of 17 STR in Fars and Isfahan.

Locus	Allele	Fars	Isfahan
DYS19	13	0.138	0.078
	14	0.404	0.472
	15	0.303	0.317
	16	0.128	0.111
	17	0.028	0.022
DYS385a	9	0.018	0.033
	10	0.028	0.017
	11	0.174	0.189
	12	0.165	0.150
	13	0.220	0.317
	14	0.165	0.150
	15	0.110	0.067
	16	0.046	0.044
	17	0.064	0.022
	18	0.009	0.011
DYS385b	11	0.018	0.011
	12	0.018	0.028
	13	0.037	0.022
	14	0.183	0.178
	15	0.156	0.167
	16	0.174	0.211
	17	0.202	0.167
	18	0.073	0.133
	19	0.083	0.039
	20	0.028	0.011
	21	0.028	0.033
DYS389I	9	0.009	0.000
	10	0.009	0.000
	12	0.165	0.150
	13	0.596	0.65

Table S2. Haploid allelic frequencies of 17 STR in Fars and Isfahan.

Locus	Allele	Fars	Isfahan
DYS389II	25	0.009	0.000
	27	0.009	0.000
	28	0.083	0.072
	29	0.294	0.417
	30	0.376	0.361
	31	0.202	0.128
	32	0.028	0.022
DYS390	20	0.000	0.006
	21	0.000	0.033
	22	0.220	0.150
	23	0.358	0.400
	24	0.248	0.278
	25	0.165	0.128
	26	0.009	0.006
DYS391	9	0.055	0.056
	10	0.743	0.706
	11	0.147	0.233
	12	0.009	0.006
	13	0.028	0.000
	14	0.018	0.000
DYS392	9	0.018	0.000
	10	0.046	0.044
	11	0.688	0.600
	12	0.046	0.050
	13	0.083	0.150
	14	0.083	0.122
	15	0.028	0.033
	16	0.009	0.000
DYS393	10	0.000	0.006
	11	0.018	0.039
	12	0.385	0.467
	13	0.459	0.389
	14	0.083	0.078
	15	0.055	0.022
DYS437	14	0.560	0.406
	15	0.266	0.450
	16	0.147	0.144
	17	0.018	0.000
	19	0.009	0.000
DYS438	9	0.312	0.378
	10	0.440	0.267
	11	0.211	0.228
	12	0.037	0.128
DYS439	10	0.193	0.139
	11	0.413	0.389
	12	0.358	0.300
	13	0.037	0.150
	14	0.000	0.022
DYS448	14	0.009	0.000
	16	0.000	0.006
	18	0.046	0.050
	19	0.275	0.378
	20	0.450	0.306
	21	0.165	0.228
	22	0.046	0.033
	23	0.009	0.000
DYS456	13	0.009	0.033
	14	0.138	0.094
	15	0.569	0.539
	16	0.257	0.217
	17	0.018	0.100
	18	0.000	0.017
	22	0.009	0.000

Table S2. Haploid allelic frequencies of 17 STR in Fars and Isfahan.

Locus	Allele	Fars	Isfahan
DYS458	14	0.055	0.028
	15	0.193	0.267
	16	0.294	0.389
	17	0.202	0.172
	18	0.138	0.094
	19	0.083	0.044
	20	0.028	0.006
	22	0.009	0.000
DYS635	19	0.000	0.006
	20	0.083	0.056
	21	0.459	0.311
	22	0.147	0.161
	23	0.193	0.333
	24	0.083	0.094
	25	0.037	0.033
	32	0.000	0.006
GATA H4	10	0.064	0.050
	11	0.468	0.456
	12	0.385	0.350
	13	0.083	0.144

Table S3. Number of alleles, gene diversity and theta in Fars and Isfahan populations, Iran.

Locus	Number of alleles[1]		Shanon index		Expected heterozygosity		Theta (θ)	
	Fars	Isfahan	Fars	Isfahan	Fars	Isfahan	Fars	Isfahan
DYS19	5(3.439)	3(2.062)	1.363	1.246	0.71577	0.66151	5.68897	3.86405
DYS385a	10(6.539)	11(6.509)	2.008	1.883	0.85491	0.81515	23.25170	14.13227
DYS385b	11(6.904)	4(3.546)	2.092	2.025	0.86306	0.85109	26.16493	22.04761
DYS389I	6(2.338)	5(3.067)	1.063	0.886	0.57764	0.51788	2.30290	1.65107
DYS389II	7(3.621)	7(3.608)	1.441	1.270	0.73055	0.67772	6.38659	4.31384
DYS390	5(3.771)	4(1.801)	1.387	1.441	0.74159	0.72688	6.98776	6.20279
DYS391	6(1.730)	6(2.481)	0.877	0.775	0.42593	0.44711	1.01717	1.13568
DYS392	8(2.031)	6(2.652)	1.167	1.250	0.51223	0.60031	1.60157	2.62986
DYS393	5(2.709)	10(5.280)	1.164	1.161	0.63677	0.62638	3.28962	3.08189
DYS437	5(2.463)	6(3.429)	1.075	1.005	0.59956	0.61558	2.61811	2.88344
DYS438	4(2.967)	5(3.527)	1.174	1.320	0.66905	0.72197	4.06491	5.96843
DYS439	4(2.968)	3(2.578)	1.172	1.372	0.66922	0.72048	4.06960	5.89966
DYS448	7(3.231)	6(2.796)	1.381	1.359	0.69691	0.71235	4.94278	5.54295
DYS456	6(2.445)	7(3.792)	1.102	1.299	0.59650	0.64593	2.57102	3.48843
DYS458	8(5.163)	8(4.048)	1.780	1.512	0.81380	0.74041	13.92080	6.91981
DYS635	6(3.520)	4(2.830)	1.490	1.578	0.72256	0.75717	5.99588	7.97937
GATA H4	4(2.643)	3(2.062)	1.105	1.155	0.62742	0.65022	3.10191	3.58670
Mean	6.294(3.440)	5.882(3.340)	1.344	1.326	0.67300	0.67500	1.77400	1.78200

[1] Values in parentheses present effective numbers of alleles.

Table S4. R_{st} values obtained for comparison between five Iranian and three regional populations. See Table S1 for complementary information about each population. The off-diagonal entries in the matrix provide the R_{st} and *P*-value between each variable pair. The values above and below the diagonal are *P*-values and R_{st} values, respectively.

Population	Fars	Isfahan	Central Iran	Mashhad	Zahedan	Pakistan	Turkey	Iraq
Fars	-	0.0449	0.0126	0.0247	0.1479	0.0000	0.0179	0.0722
Isfahan	0.0079	-	0.7718	0.3732	0.1745	0.0000	0.0492	0.0631
Central Iran	0.0130	-0.0024	-	0.3669	0.1372	0.0000	0.0034	0.0447
Mashhad	0.0117	0.0002	0.0004	-	0.2293	0.0000	0.0410	0.1590
Zahedan	0.0046	0.0031	0.0044	0.0027	-	0.0000	0.0131	0.0720
Pakistan	0.0856	0.0773	0.0783	0.0462	0.0692	-	0.0000	0.0000
Turkey	0.0085	0.0044	0.0104	0.0057	0.0108	0.0581	-	0.5088
Iraq	0.0085	0.0071	0.0093	0.0041	0.0087	0.0518	0.0008	-

Table S5. The F_{st} values obtained for comparison between five Iranian and three regional populations (See Table S1 for complementary information about each population). The off-diagonal entries in the matrix provide the F_{st} and *P*-value between each variable pair. The values above and below the diagonal are *P*-values and F_{st} values, respectively.

Population	Fars	Isfahan	Central Iran	Mashhad	Zahedan	Pakistan	Turkey	Iraq
Fars[1]	-	0.0017	0.0010	0.0812	0.1256	0.0001	0.0003	0.0616
Isfahan[2]	0.0008	-	1.0000	0.0079	0.0513	0.0000	0.0000	0.0050
Central Iran	0.0011	-0.0033	-	0.0096	0.0244	0.0000	0.0000	0.0032
Mashhad[3]	0.0002	0.0006	0.0007	-	1.0000	0.0000	1.0000	0.1893
Zahedan[4]	0.0003	0.0005	0.0007	0.0000	-	0.0000	1.0000	0.4948
Pakistan[5]	0.0052	0.0056	0.0058	0.0048	0.0049	-	0.0000	0.0002
Turkey[6]	0.0003	0.0007	0.0009	0.0000	0.0000	0.0049	-	0.0606
Iraq[7]	0.0004	0.0007	0.0011	0.0001	0.0000	0.0050	0.0001	-