

**A preliminary report on X-chromosome sequence variability within Iranian population:
Reporting new potential SNPs/ sequence variants as a source of population genetic markers**

MASOUD SHEIDAI^{1*}, MEHDI HEYDARI¹, GOLNAZ ATRI¹, NASER KALHOR²,
FAHIMEH KOOHDAR¹, ZAHRA NOORMOHAMMADI³

*¹Faculty of Life Sciences and Biotechnology, Shahid Beheshti University, Tehran, Iran; Email:
msheidai@yahoo.com*

*²Department of mesenchymal stem cell, Academic Center for Education, Culture and Research,
Qom branch, IRAN*

*³Department of Biology, School of Basic Sciences, Science and Research Branch, Islamic Azad
University, Tehran, Iran*

**This article has been accepted for publication and undergone full peer review but has not
been through the copyediting, typesetting, pagination and proofreading process, which may
lead to differences between this version and the Version of Record.**

Please cite this article as:

**Masoud Shedai, Mehdi Heydari, Golnaz Atri, Naser Kalhor, Fahimeh Koohdar, Zahra
Noormohammadi (2020). A preliminary report on X-chromosome sequence variability
within Iranian population: Reporting new potential SNPs/ sequence variants as a source of
population genetic markers. Caryologia, Just Accepted.**

Abstract

X-chromosome and its independent loci, recently has gained importance in genetic investigations that are concerned with diseases control-case studies, as well as in population genetic analyses and phylogenetic investigations. The present study was performed as a preliminary investigation on X-chromosome sequence variability as a source of genetic diversity within Iranian population and to investigate potential use of these sequences for differentiating the ethnic groups within the country. A limited random sample could be obtained from four different ethnic populations, which were sequenced for TEX-11 gene of the X-chromosome. At the beginning we used seven specific primers related to seven known SNPs within TEX-11 gene. Later on we searched for the other known as well as new potential SNPs or sequence variants in the studied samples. We could identify 87 potential SNPs or sequence variants. Genetic diversity analysis revealed the presence of five haplotype groups within the studied sequences. AMOVA revealed no significant genetic differentiation among sample driven from the ethnic populations. Moreover, Mantel test did not show significant association between geographical distance and genetic distance of these samples, indicating that gene flow occurred among these populations with no geographical obstacle. This is the first report on X-chromosome sequence diversity analysis in Iran.

Key word. X-chromosome, SNPs, population, haplotype

INTRODUCTION

Population genetic study produces data on genetic structure and genetic diversity within target populations and investigates genetic admixture and gene flow versus reproductive isolation among the studied populations. It also may identify specific genotypes or genes in local geographical or ethnic populations (Freeland et al. 2011).

X-chromosome markers are of potential use in population and forensic genetic studies (Zhang et al. 2015), due to their peculiar transmission property (Garrigan et al. 2007). These genetic markers are transmitted between both sexes in each generation, and they may portraiture a different picture from uniparental genomes. These molecular markers are sensitive to the evolutionary processes like population substructure or fragmentation and genetic drift, as their effective population size is reduced in relation to autosomes (Garrigan et al. 2007; Zhang et al. 2015).

Nowadays, the number of genetic investigators interested in applying X-chromosome markers in genetic studies grows as these markers show higher efficiency parameters than autosomes in special kinship investigations involving mainly female offspring. These genetic X chromosome markers are X-STRs, X-SNPs and more recently X-Indels, that also may be utilized in studying the genetic structure of human populations, and investigating the ancestry proportions in the admixed populations as well as for forensic investigations (Garrigan et al. 2007; Zhang et al. 2015).

The present investigation was carried out as a side work during investigation on TEX-11 gene and its association with infertility. Therefore, we also used TEX-11 sequences as a population genetic marker to produce preliminary data on sequence variability and gene flow in a small

sample size of Iranians who were randomly chosen from four ethnic populations. This is the first report on X-chromosome sequence diversity.

Material and Methods

Samples studied

In total 30 persons, were randomly selected for sequencing. The studied samples were taken from four ethnic Iranian populations, 1- Turk, 2- Fars, 3- Lor, and 4- Kurd.

Gene and SNPs selection

TEX11 is one the most important genes in the spermatogenesis process that has tissue specific expression. It is expressed in pancreas and testis only. The SNP selection was based on PHYRE2 online software and by determining the rate of amino acids conservation and protein structure sensitivity to amino acid alteration (Yang et al. 2015). The SNPs chosen were not studied before.

DNA extraction and PCR details

The blood samples were obtained from Jihad center of Qom University (informed consent was obtained from all participate). The standard salting out method was used for extraction of genomic DNA from blood samples. The quality of DNA samples were examined by 1% agarose gel electrophorese. Target TEX11 gene was amplified by PCR reaction, while forward as well as reverse primers were designed by oligo7 version 7.56 (Table1).

Table1. Seven different TEX-11 gene regions for sequencing

		Primers information	
Forward	rs775667438	5' GCATGGCATCTATCTCTCTG 3'	TM= 57
Reverse		5' GGTGAATTATGGGTGTTCTC 3'	TM= 57
Forward	rs746940663	5' TGCCCACGATACCTACTG 3'	TM= 54
Reverse		5' GACTGAATATGGACAGAGGA 3'	TM= 54
Forward	rs6525433	5' AAATGCTAACTGTTGCTTTT 3'	TM= 55
Reverse		5' CCCACGATACCTACTGACTC 3'	TM= 53
Forward	rs756553436	5' CAACATCAAGGTGCTCGCAG 3'	TM= 57
Reverse		5' GAAGATGCCTGTCACTGTGG 3'	TM= 58
Forward	rs867296518	5'ATCAGCGATGACATTTCCCTAC 3'	TM= 57
Reverse		5' GAGAGGGAGACAATAGACCGAG 3'	TM= 57
Forward	rs977597709	5' CCCAATTTGTGGGATGTGGACAG 3'	TM= 60
Reverse		5' TCTGTTGGGTTTCATTTCTGACG 3'	TM= 59
Forward	rs1056191384	5' CTTGTTCAAAGGTACACAGC 3'	TM= 54
Reverse		5' CATGTAACCTCACTGGATCTCG 3'	TM= 54

PCR was done by thermo cycler system (Genetix Biotech, Australia), having the initial denaturation of 95°C for 4 min. This was then followed by 40 cycles of denaturation at 95°C for 30 s, and annealing at different temperatures according to different primers used (Table1) for 30 s, and extension at 72°C for 45 s. The final extension was done at 72°C for 10 min.

PCR reaction was carried out in total volume 25 μ l containing 3 μ l of DNA samples, 22 μ l Master Mix (17.2 μ l H₂O, 2 μ l Buffer 10X, 0.3 μ l MgCl₂ 100 mM, 0.4 μ l dNTP-Mix 40 mM, 0.8 μ l Forward primer 14.72 ng/ μ l and 0.8 μ l Reverse primer 12.77 ng/ μ l) and 0.5 μ l Taq polymerase (5 u/ μ l). The PCR fragments were separated by 1% agarose gel electrophoresis and visualized with green viewer staining (5u/ μ l).

Data analyses

Sequence alignment and curation was done by MUSCLE program implemented in MEGA 7 software (Tamura et al. 2012). Multidimensional scaling methods (MDS) was performed on these sequences to investigate the ethnic populations genetic differentiation. This can be taken as indicating gene flow and genetic admixture among ethnic groups. For this, Kimura 2-parameters were used as genetic distance of the studied samples.

Haplotype groups and haplotype diversity were determined by TCS networking as performed in POPART (Population Analysis with Reticulate Trees) program (<http://popart.otago.ac.nz>). Significant genetic difference among the ethnic populations was studied by analysis of molecular variance (AMOVA) and by calculation PHIST as implemented in POPART ver. 3 (Librado and Rozas 2009). The Mantel test was used to investigate association between genetic distance and geographical distance of the studied ethnic groups (Podani 2000).

RESULTS

Sequence variability and haplotype groups

In total we obtained 1935 bp length DNA after alignment and curation. The nucleotide diversity obtained was $\pi = 0.014$, the number of segregating sites = 206, the No. of parsimony-informative sites was 6, and Tajima's D statistic was $D = -1.92938$ ($p(D \geq -1.92938) = 0.984141$). All these

results indicated a low degree of nucleotide substitution and that these substitutions are not under selective pressure.

The studied samples differed in sequences as they formed different clusters/ clades in TCS network (Fig. 1). Some of the sequences were placed close to the out-group (main sequence = No. 1 in Fig.1), and formed the haplotype group A, due to sequence similarity. However, the samples within this haplotype group, also revealed some degree of sequence variability.

The samples within B and C haplotype groups were placed close to the A group. Members of these haplotype groups also differed in the number of nucleotide substitutions. However, haplotype groups D and E were placed far from the A group, and therefore, are much more deviant from the reference sequence.

The highest number of nucleotide substitutions occurred in the sequence 31 in haplotype group D (30 substitutions), followed by sequence 23 of the haplotype group E (23 substitutions).

Gene flow and population admixture

We investigated the population genetic parameters based on ethnic populations from which our samples were collected (Table2 and 3).

Table 2. Genetic diversity parameters determined in the studied SNPs

	No. Haplotypes	Haplotype diversity	Nucleotide diversity	Polymorphic Sites
SNP1	8	0.71	0.04	287
SNP2	7	0.40	0.03	68
SNP3	5	0.30	0.03	37
SNP4	2	0.01	0.005	3
SNP5	6	0.50	0.03	40
SNP6	8	0.70	0.04	250
SNP7	2	0.02	0.006	10

Table 3. Details of ethnic populations in which SNP sequences were potentially associated with the male infertility

Number	SNP	Fars	Lor	Kurd	Turk
SNP1	Rs6525433	-	+	-	-
SNP2	Rs1056191384	-	+	+	-
SNP3	Rs746940663	+	-	-	-
SNP4	Rs977597709	+	+	-	-
SNP5	Rs775667438	+	-	+	+
SNP6	Rs867296518	+	-	-	+
SNP7	Rs756553436	-	-	-	-

AMOVA performed did not produced significant difference among the studied ethnic populations ($F_{\text{st}} = 0.2$, $P = 0.26$). AMOVA revealed that only 3% of total genetic variation occurred among the studied ethnic populations, while 97% was due to within population genetic variability. Therefore, it indicates that these ethnic populations are not genetically differentiated and that they have/had frequent gene flow. This is also supported by MDS plot based on ethnic groups (Fig. 2) which also reveals high degree of genetic admixture among the studied ethnic populations.

The Mantel test did not produce significant association between genetic distance and geographical distance of ethnic populations studied ($r = 0.10$, $P = 0.3$). Therefore, geographical distance was not an obstacle for migration and gene flow for the studied populations.

New potential SNPs / sequence variants within TEX-11 gene

In total we found 87 SNPs and sequence variants within TEX-11, details of which are provided in Table 4. Out of 87 SNPs observed, 18 SNPs were previously reported as RS in NCBI (Table 4), while the rest of them are new reports. However, at present we consider these SNPs as polymorphism within Iranian samples studied.

Table4. Details of SNPs/sequence polymorphism with TEXT-11 gene in the studied samples and NCBI.

NO	Position in sequence	Alleles	Reference SNP	N O	Position in sequence	Alleles	Reference SNP
1	49	T/A	-	81	1615	T/G	-
2	52	A/T		82	1617	T/C	-
3	59	A/G		83	1620	C/T	-
4	95	G/A	RS181582479	84	1622	G/T	-
5	442	G/T	-	85	1624	T/C	-
6	485	A/C/G	-	86	1627	C/T	-
7	488	G/T/A	--	87	1631	G/A	-
8	490	A/T/A	-	88	1632	A/G	-
9	491	C/T(A)	RS375000923	89	1635	G/A	-
10	492	T/A/C	-	90	1636	T/G	--
11	493	C/T/A	-	91	1638	T/C	-
12	494	T/G	-	92	1640	T/C	-
13	495	G/A(T)	RS755626796	93	1642	A/C	-
14	496	A/C/G	-	94	1651	T/A	-
15	497	A/C/T	-	95	1653	G/A	-
16	502	T/G/C	-	96	1654	T/G	-
17	505	C/T(A)	RS1296061122	97	1655	T/G	-
18	506	T/C(A)	RS1247146060	98	1661	T/C	-
19	507	C/A/T	-	99	1665	A/C	-
20	508	T/C	-	100	1666	C/T	RS1186389277
21	509	C/T(A)	RS1365016323	101	1678	G/T	-
22	512	C/G/A	-	102	1682	G/A	-
23	513	A/G	RS748927377	103	1691	A/G	-
24	514	G/C	-	104	1695	G/A	-
25	515	T/C(G)	RS1341488009	105	1697	G/T	-
26	516	A/C/T	-	106	1699	T/G	-
27	521	T/C	-	107	1700	G/C	-

28	525	C/T	-	10	1703	A/C	-
				8			
29	534	A/C	-	10	1708	T/G	-
				9			
30	620	A/C	-	11	1710	A/C	-
				0			
31	646	T/A	-	11	1711	A/C	-
				1			
32	700	A/G	-	11	1712	A/G	-
				2			
33	721	A/T	-	11	1713	G/C	-
				3			
34	752	T/C	-	11	1714	A/C	-
				4			
35	766	A/G	-	11	1717	A/T	-
				5			
36	811	C/T	RS181582479	11	1721	G/C/A	-
				6			
37	867	C/A/T	-	11	1729	G/C	-
				7			
38	892	G/T	-	11	1730	A/G	RS99398839
				8			8
39	906	C/G	-	11	1731	A/G	-
				9			
40	908	T/C	-	12	1739	A/G	-
				0			
41	953	G/A	-	12	1744	A/G	RS11389495
				1			7
42	957	T/G	-	12	1750	A/C	-
				2			
43	974	A/C	-	12	1754	A/T	-
				3			
44	982	T/A	-	12	1760	T/A	-
				4			
45	1032	T/C	-	12	1764	G/C	-
				5			
46	1041	G/A	-	12	1768	T/A	-
				6			
47	1068	A/G	Rs774596712	12	1774	T/A	-
				7			
48	1070	T/C	-	12	1780	C/T	-
				8			
49	1071	C/G	-	12	1781	T/C	-
				9			
50	1074	C/A	-	13	1782	G/A	-
				0			

51	1144	A/C	-	13	1788	T/C	-
52	1412	C/T	-	13	1789	A/C/T	-
53	1413	T/G	-	13	1826	T/A	-
54	1414	G/T	-	13	1827	G/C	-
55	1429	A/G	-	13	2332	C/T	-
56	1430	G/A	RS1188518295	13	2335	T/A	-
57	1431	A/G	-	13	2340	T/C	-
58	1441	C/T	-	13	2343	A/T	-
59	1444	A/T	-	13	2353	T/A	-
60	1451	G/A	-	14	2357	G/A	-
61	1456	A/G	-	14	2359	G/A	-
62	1460	T/G	-	14	2360	G/A	-
63	1461	C/T	-	14	2364	T/C	-
64	1486	C/A	-	14	2374	T/A	--
65	1488	C/T(G)	RS1161677215	14	2399	G/C	-
66	1489	A/G	-	14	2412	C/T	-
67	1495	A/T/C	-	14	2414	A/T	-
68	1496	G/C	-	14	2417	T/A	-
69	1497	G/C/T	-	14	2421	G/C	-
70	1499	A/T	-	15	2430	A/T	-
71	1501	C/T	-	15	2435	T/A	-
72	1599	T/C	-	15	2436	A/T	-
73	1600	T/A/C	-	15	2445	C/T	-

74	1601	T/G/A	-	15 4	2454	C/A	-
75	1603	A/G	-	15 5	2455	A/T	RS12567792 49
76	1605	C/T	-	15 6	2474	G/C	RS99348824 5
77	1607	C/A/T	-	15 7	2551	T/A	-
78	1608	A/T	-	15 8	2556	C/G	-
79	1612	A/G/T	-	15 9	2644	C/T	-
80	1614	G/T(A)	RS1452980045	16 0	2645	G/T	-

Details of the SNP nucleotides are also provided for each sequence position. In some cases multiple substitutions were observed within the studied samples (Fig. 3).

DISCUSSION

We observed almost a moderate level of sequence variability in X-chromosome. In a similar investigation, Nachman et al. (1998), sequenced 11,365 bp from introns of seven X-linked genes in 10 humans, one chimpanzee, and one orangutan and reported the average value for pi as low as 0.063% with Standard error = 0.036%. These authors reported a positive correlation between heterozygosity and rate of recombination. In total it was suggested that the joint effects of selection and linkage are important in shaping patterns of nucleotide variation in humans (Nachman et al. 1998).

The present study revealed some degree of genetic admixture and gene flow among the ethnic populations studied. In a similar study Zhang et al. (2015), used 34 X-Chromosome markers (18 X-STRs and 16 X-Indels), to investigate genetic variability and admixture in ethnic

populations of China. They reported genetic variability of 0.4-0.7 for the studied populations, while Phylogenetic tree and PCA analyses revealed a clear pattern of population differentiation. The study suggested that geographic isolation and interactions play significant roles in differentiation of genetic constitution of ethnic groups.

The use of the X chromosome in population genetics is still in its infancy. It has already proved its worth in studies of the early history of modern Homo sapiens (Harris and Hey 1999), but in most research areas its potential remains largely untapped. Haplotypes obtained based on coding as well as non-coding sequences in X-chromosome are of genetic analysis importance, particularly in genetic variability analysis (Kaessmann et al. 1999), as also revealed in present study

In conclusion, the present study revealed potential use of coding as well as non-coding sequences in genetic variability studies.

References

- Freeland JR, Kirk H, Peterson SD .2011. Molecular Ecology.In: Wiley-Blackwell, U.K. (2nded), 449.
- Garrigan D, Kingan S B, Pilkington M M, Wilder JA, Cox M P, Soodyall H, Strassmann B, Destro-Bisol G, de Knijff P, Novelletto A, Friedlaender J, Hammer M F. 2007. Inferring Human Population Sizes, Divergence Times and Rates of Gene Flow from Mitochondrial, X and Y Chromosome Resequencing Data. *Genetics* 177: 2195–2207.
- Harris E E, Hey JX .1999. chromosome evidence for ancient human histories. *Proc. Natl Acad. Sci. USA* 96, 3320–3324 A study of haplotypes in a 4.2-kb X-chromosome locus, showing surprisingly large differences in variation between African and non-African human populations.
- Kaessmann H, Heissig F, von Haeseler A, Pääbo S. 1999. DNA sequence variation in a non-coding region of low recombination on the human X chromosome. *Nature Genetics* 22:78–81.
- Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
- Nachman MW, Bauer VL, Crowel SL, Aquadro CF. 1998. DNA variability and recombination rates at X-linked loci in humans. *Genetics* 150: 1133-1141.
- Podani J. 2000. Introduction to the Exploration of Multivariate Data. Backhuyes: Leiden, p. 407.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2012. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 28:2731e2739.

Yang F, Silber Sh, Leu NA, Oates RD, Marszalek JD, Skaletsky H, Brown LG, Rozen S, Page DC, Wang PJ. 2015. TEX11 is mutated in infertile men with azoospermia and regulates genome-wide recombination rates in mouse. *EMBO Molecular Medicine* 7: 1198–1210

Zhang S, Bian Y, Li L, Sun K, Wang Z, Zhao Q, Zha L, Cai J, Gao Y, Ji C, Li C. 2015. Population genetic study of 34 X-Chromosome markers in 5 main ethnic groups of China. *Scientific Reports* 5:17711.

Zhang X, Ding M, Ding X, Li T, Chen H. 2015. Six polymorphisms in genes involved in DNA double-strand break repair and chromosome synapsis: association with male infertility. *Systems Biology in Reproductive Medicine* 61(4): 187–193.

Accepted Manuscript

Figure1. TCS network of the studied sequences showing five major haplotype groups.

Figure2. MDS plot of the studied sequences based on ethnic group samples showing admixture of the samples.

Figure3. SNP diversity within TEXT-11 gene in the studied samples.

Accepted Manuscript