Caryologia
International Journal of Cytology, Cytosystematics and Cytogenetics

Firenze University Press
www.fupress.com/caryologia

Citation: M. Sheidai, M. Heydari, G. Atri, N. Kalhor, F. Koohdar, Z. 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* 73(4): 77-83. doi: 10.13128/caryologia-551

Received: July 16, 2019

Accepted: September 07, 2020

Published: May 19, 2021

Copyright: © 2020 M. Sheidai, M. Heydari, G. Atri, N. Kalhor, F. Koohdar, Z. Noormohammadi. This is an open access, peer-reviewed article published by Firenze University Press (http://www.fupress.com/caryologia) and distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

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 Heydari[1], Golnaz Atri[1], Naser Kalhor[2], Fahimeh Koohdar[1], Zahra Noormohammadi[3]

[1] *Faculty of Life Sciences and Biotechnology, Shahid Beheshti University, Tehran, Iran*
[2] *Department of mesanchymal stem cell, Academic Center for Education, Culture and Research, Qom branch, Iran*
[3] *Department of Biology, School of Basic Sciences, Science and Research Branch, Islamic Azad University, Tehran, Iran*
*Corresponding author. E-mail: msheidai@yahoo.com

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.

Keyword: 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 portrait 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 (Table 1).

PCR was done by thermo cycler system (Genetix Biotech, Australia), having the initial denaturation of

Table 1. 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’ TCTGTTGGGTTTCATTTCCTGACG 3’	TM= 59
Forward	rs1056191384	5’ CTTGTTCAAAGGTACACAGC 3’	TM= 54
Reverse		5’ CATGTAACTTCACTGGATCTCG 3’	TM= 54

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 µl containing 3 µl of DNA samples, 22 µl Master Mix (17.2 µl H2O, 2 µl Buffer 10X, 0.3 µl MgCl2 100 mM, 0.4 µl dNTP-Mix 40 mM, 0.8 µl Forward primer 14.72 ng/µl and 0.8 µl Reverse primer 12.77 ng/µl) and 0.5 µl Taqpolymerase (5 u/µl).The PCR fragments were separated by 1% agarose gel electrophoresis and visualized with green viewer staining (5u/µl).

Data analyses

Sequence alignment and curation was done by MUSCLE program implemented in MEGA 7 software (Tamura et al. 2012). Multidimentional 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 POP-ART (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 >= -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 outgroup (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.

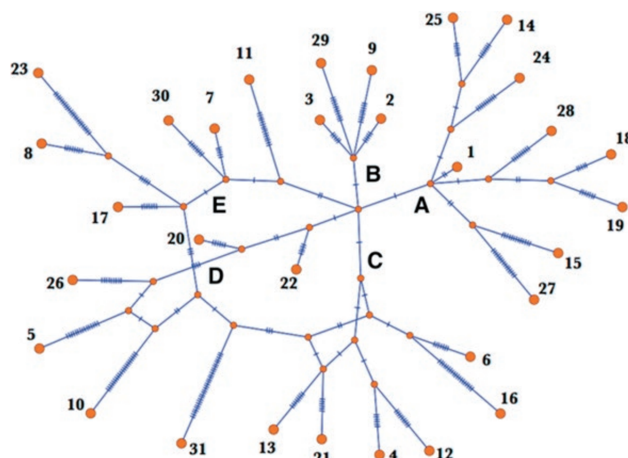

Figure 1. TCS network of the studied sequences showing five major haplotype groups.

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 (Table 2 and 3).

AMOVA performed did not produced significant difference among the studied ethnic populations (Phist = 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 with 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).

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

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

Therefore, geographical distance was not an obstacle for migration and gene flow for the studied populations.

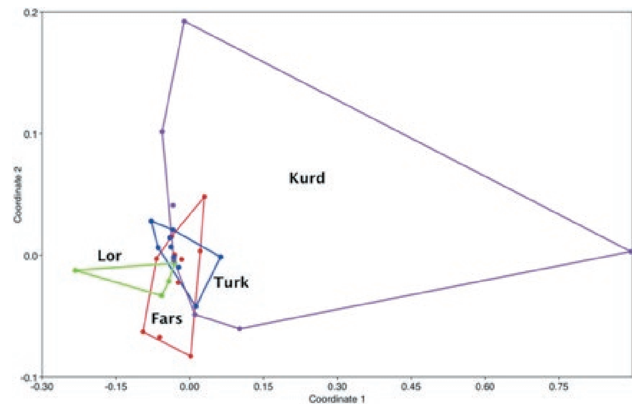

Figure 2. MDS plot of the studied sequences based on ethnic group samples showing admixture of the samples.

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.

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).

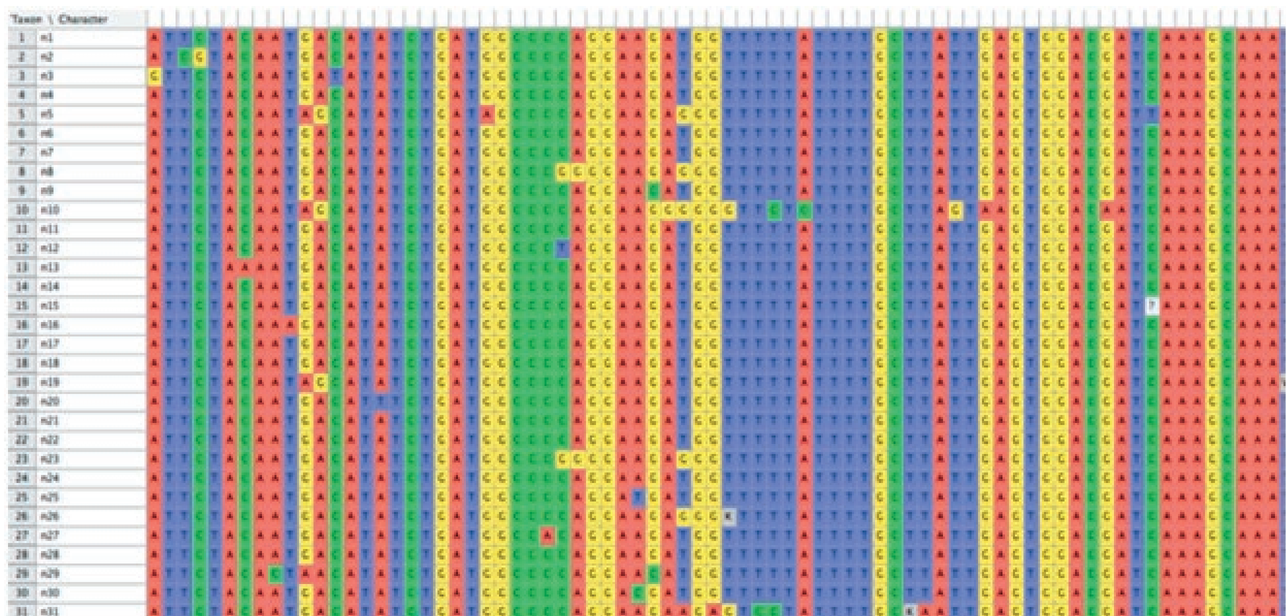

Figure 3. SNP diversity within TEXT-11 gene in the studied samples.

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

Reference SNP	Alleles	Position in sequence	NO	Reference SNP	Alleles	Position in sequence	NO
-	T/G	1615	81	-	T/A	49	1
-	T/C	1617	82		A/T	52	2
-	C/T	1620	83		A/G	59	3
-	G/T	1622	84	RS181582479	G/A	95	4
-	T/C	1624	85	-	G/T	442	5
-	C/T	1627	86	-	A/C/G	485	6
-	G/A	1631	87	--	G/T/A	488	7
-	A/G	1632	88	-	A/T/A	490	8
-	G/A	1635	89	RS375000923	C/T(A)	491	9
--	T/G	1636	90	-	T/A/C	492	10
-	T/C	1638	91	-	C/T/A	493	11
-	T/C	1640	92	-	T/G	494	12
-	A/C	1642	93	RS755626796	G/A(T)	495	13
-	T/A	1651	94	-	A/C/G	496	14
-	G/A	1653	95	-	A/C/T	497	15
-	T/G	1654	96	-	T/G/C	502	16
-	T/G	1655	97	RS1296061122	C/T(A)	505	17
-	T/C	1661	98	RS1247146060	T/C(A)	506	18
-	A/C	1665	99	-	C/A/T	507	19
RS1186389277	C/T	1666	100	-	T/C	508	20
-	G/T	1678	101	RS1365016323	C/T(A)	509	21
-	G/A	1682	102	-	C/G/A	512	22
-	A/G	1691	103	RS748927377	A/G	513	23
-	G/A	1695	104	-	G/C	514	24
-	G/T	1697	105	RS1341488009	T/C(G)	515	25
-	T/G	1699	106	-	A/C/T	516	26
-	G/C	1700	107	-	T/C	521	27
-	A/C	1703	108	-	C/T	525	28
-	T/G	1708	109	-	A/C	534	29
-	A/C	1710	110	-	A/C	620	30
-	A/C	1711	111	-	T/A	646	31
-	A/G	1712	112	-	A/G	700	32
-	G/C	1713	113	-	A/T	721	33
-	A/C	1714	114	-	T/C	752	34
-	A/T	1717	115	-	A/G	766	35
-	G/C/A	1721	116	RS181582479	C/T	811	36
-	G/C	1729	117	-	C/A/T	867	37
RS993988398	A/G	1730	118	-	G/T	892	38
-	A/G	1731	119	-	C/G	906	39
-	A/G	1739	120	-	T/C	908	40
RS113894957	A/G	1744	121	-	G/A	953	41
-	A/C	1750	122	-	T/G	957	42
-	A/T	1754	123	-	A/C	974	43
-	T/A	1760	124	-	T/A	982	44
-	G/C	1764	125	-	T/C	1032	45
-	T/A	1768	126	-	G/A	1041	46
-	T/A	1774	127	Rs774596712	A/G	1068	47
-	C/T	1780	128	-	T/C	1070	48
-	T/C	1781	129	-	C/G	1071	49

Reference SNP	Alleles	Position in sequence	NO	Reference SNP	Alleles	Position in sequence	NO
-	G/A	1782	130	-	C/A	1074	50
-	T/C	1788	131	-	A/C	1144	51
-	A/C/T	1789	132	-	C/T	1412	52
-	T/A	1826	133	-	T/G	1413	53
-	G/C	1827	134	-	G/T	1414	54
-	C/T	2332	135	-	A/G	1429	55
-	T/A	2335	136	RS1188518295	G/A	1430	56
-	T/C	2340	137	-	A/G	1431	57
-	A/T	2343	138	-	C/T	1441	58
-	T/A	2353	139	-	A/T	1444	59
-	G/A	2357	140	-	G/A	1451	60
-	G/A	2359	141	-	A/G	1456	61
-	G/A	2360	142	-	T/G	1460	62
-	T/C	2364	143	-	C/T	1461	63
--	T/A	2374	144	-	C/A	1486	64
-	G/C	2399	145	RS1161677215	C/T(G)	1488	65
-	C/T	2412	146	-	A/G	1489	66
-	A/T	2414	147	-	A/T/C	1495	67
-	T/A	2417	148	-	G/C	1496	68
-	G/C	2421	149	-	G/C/T	1497	69
-	A/T	2430	150	-	A/T	1499	70
-	T/A	2435	151	-	C/T	1501	71
-	A/T	2436	152	-	T/C	1599	72
-	C/T	2445	153	-	T/A/C	1600	73
-	C/A	2454	154	-	T/G/A	1601	74
RS1256779249	A/T	2455	155	-	A/G	1603	75
RS993488245	G/C	2474	156	-	C/T	1605	76
-	T/A	2551	157	-	C/A/T	1607	77
-	C/G	2556	158	-	A/T	1608	78
-	C/T	2644	159	-	A/G/T	1612	79
-	G/T	2645	160	RS1452980045	G/T(A)	1614	80

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.