UDC 575.
https://doi.org/10.2298/GENSR23010229S
Original scientific article

TEXT-11 SNPS AND NEIGHBORING SEQUENCES ROLE IN CAUSING MALE INFERTILITY IN SOME ETHNIC GROUPS OF IRANIAN AND THEIR POTENTIAL ROLE IN ESTIMATING DIVERGENCE TIME

Masoud SHEIDAI[1*], Sogol ALLAHVERDI[2], Narges ANVARI[2], Somayeh EIVAZI KHAMENE[2], Mozhgan GHANDEHARI ALAVIJEH[2], Marziyeh SEYED GHOREISHI[2], Fatemeh JAMSHIDI[2], Ghazal KHOSRAVANI[2], Naser KALHOR[3], Fahimeh KOOHDAR[1]

[1]Faculty of Life Sciences and Biotechnology, Shahid Beheshti University, Tehran, Iran
[2]Department of Genetics, Faculty of Advanced Science and Technology, Tehran Medical Sciences, Islamic Azad University, Tehran, Iran
[3]Department of mesanchymal stem cell, Academic Center for Education, Culture and Research Qom branch, Iran

Sheidai M., S. Allahverdi, N. Anvari, S. Eyvazi Khamene, M. Ghandehari, M. Ghoreshi, F. Jamshidi, G. Khosravani, N. Kalhor, F. Koohdar (2023). *Text-11 snps and neighboring sequences role in causing male infertility in some ethnic groups of Iranian and their potential role in estimating divergence time*- Genetika, Vol 55, No.1, 229-244.

X-chromosome and its independent loci recently gained importance in genetic investigations concerned with disease control-case studies, population genetic analyses, and phylogenetic investigations. Infertility and lack of normal reproduction are important issues in health. Aazoospermia is a kind of infertility in men whose semen contains no sperm and can affect about 1-20% of the male population. The majority of cases of spermatogenic failure in humans are idiopathic, and the underlying causes are postulated to be genetic. We analyze the potential association between seven X-Chromosome SNPs and the neighboring sequences with azoospermia in four ethnic populations in Iran. We

Corresponding author: Masoud Sheidai, Faculty of Life Sciences and Biotechnology, Shahid Beheshti University, Tehran, Iran, Email: msheidai@yahoo.com

also carried out population genetic analyses to produce data on genetic variability, population structure, gene flow, and time of divergence in these ethnic groups, for the first time. As the X-chromosome genes/ SNPs are in homozygous condition, for the association study we used multivariate statistical methods, which revealed the potential role of variable sites close to the studied SNPs in male infertility. Haplotype groups were identified for each of the seven SNPs. The results showed gene flow among ethnic groups due to migration and inter-breeding and Bayesian analysis of sequence data suggested divergence time for ethnic populations to be between 2-30 KY.

Keywords: Azoospermia, Population, SNPs, X-chromosome

INTRODUCTION

Infertility is one of the reproductive health problems, affecting males and females. It is defined as the inability to conceive after a prolonged period (HULL *et al.,* 1985; GHIEH *et al.,* 2019). Infertility may result due to various causes such as physiological, environmental, social, and genetic factors. Recent studies indicate that different molecular and genetic pathways regulate fertility and that hundreds of genes operate in these pathways (MATZUK and DOLORES, 2008; SALAMONSEN *et al.,* 2013).

Azoospermia is a state of infertility in which men's ejaculate contains no sperm.In humans, azoospermia affects about 1% of the male population and may be seen in up to 20% of male infertility situations (JARVIS *et al.,* 2010). Known genetic causes of azoospermia in humans include Y chromosome deletion and chromosomal abnormalities such as Klinefelter syndrome (47, XXY); these account for ~25% of spermatogenic failure in otherwise healthy men (REIJO *et al.,* 1995; 1996; VAN ASSCHE *et al.,* 1996). Therefore, the majorities (~75%) of cases of spermatogenic failure in humans are idiopathic, and the underlying causes are postulated to be genetic.

Three X-linked genes (Tex11, Taft l, and Nxf2) are known to regulate the male fertility (YANG *et al.,* 2008; ZHENG *et al.,* 2010; YANG *et al.,* 2015). Mutations in these X-linked genes can bring about infertility in men as the males are hemizygous for the X chromosome, and therefore, mutations in single-copy X-linked genes cannot be compensated by a corresponding wild-type allele such as in heterozygous carriers of autosomal recessive mutations.

Recent studies revealed that along with the sequences known to be promoting or regulating transcription and translation, there appears to be a large set of functional sequence ‘dark matter’, whose importance is yet to be understood (PONTING and LUNTER, 2006; TSAGAKIS *et al.,* 2020). It is estimated that between 10 and 15% of patients with rare Mendelian phenotypes exhibit no changes to a gene's coding sequence, despite incontrovertible evidence of its association with disease (EMISON *et al*., 2005). Therefore, it must be assumed that the as-yet-unknown mutations lie in the functional non-coding portions of the human genome. Indeed, mutations in intronic elements (EMISON *et al*., 2005), promoters and untranslated regions (UTRs) have, on occasion, been associated with disease (PONTING and LUNTER, 2006).

Tex11 is essential for male fertility. Disruption of Tex11 gene function causes meiotic arrest in mouse (YANG *et al.,* 2008) and has potential importance in human spermatogenesis (YANG *et al.,* 2015). YANG *et al.* (2015) show that, only three TEX11 mutations (frameshift mutation in exon 16, splice site mutation in intron 21, and V748A missense mutation) indicates

an infertility-causing mutation frequency in human TEX11 of ~1%. Given that hundreds, if not thousands, of genes specifically regulate fertility, finding a causative mutation frequency of 1% in TEX11 is highly significant (YANG *et al.,* 2015).

ZHANG *et al.* (2015a), investigated four genes involved in DNA double-strand break repair and chromosome synapsis, i.e., testis expressed gene 11 (TEX11), testis expressed gene 15 (TEX15), mutL homolog 1 (MLH1), and homolog 3 (MLH3), that play critical roles in genome integrity, meiotic recombination, and gametogenesis. They reported the possible association between single nucleotide polymorphisms (SNPs) in these genes and idiopathic male infertility involving azoospermia or oligozoospermia.

Along with genetic disease association studies, X-chromosome markers are of potential use in population genetic studies (ZHANG *et al.,* 2015b). X-Chromosome is gaining significant importance in population and forensic genetic studies due to the special transmission property. The X-Chromosome is transmitted between both sexes in each generation, telling a different story from uniparental genomes. Moreover, its effective population size is reduced in relation to autosomes, making it more sensitive to the effects of population substructure and genetic drift. And X-Chromosome markers show higher efficiency parameters than autosomes in special kinship investigations involving mainly female offspring, making them suitable for forensic application. Thus, growing number of scientists is becoming interested in X-Chromosome research, using X markers (X-STRs, X-SNPs and more recently X-Indels) for studying the genetic structure of human populations, ancestry proportions in admixed populations as well as for forensic investigations (ZHANG *et al.,* 2015b).

In the present study we investigated potential association between seven Text-11 gene SNPs and the neighboring sequences with azoospermia in four ethnic populations of Iran for the first time. We also tried to estimate the probable divergence time for the studied ethnic groups and also identify their route of migration within the country. We used both RFLP and sequencing approaches.

MATERIAL AND METHODS

DNA samples

In this study, 100 fertile men and 60 infertile men with azoospermia were recruited Jihad of Qom University (informed consent was obtained from all participants). All patients with abnormal karyotypes and Y chromosome microdeletions were excluded from the study. Semen analysis based on the World Health Organization criteria was performed for all participants (COOPER *et al*., 2010). The standard salting-out method was used for the extraction of genomic DNA from blood samples. The quality of DNA samples was examined by 1% agarose gel electrophorese (MILLER *et al.,* 1988).

TEX11 is one the most important genes during the spermatogenesis process and this gene has tissue specific expression and it expresses in pancreas and testis only. Several previous studies have reported the TEX11 importance in spermatogenesis (SHEIDAI *et al.,* 2021) and also we approved the role, importance and specification of this gene in testis and spermatogenesis based on GTX database (https://gtexportal.org/home/). All of the SNPs (7 SNPs) are novel. 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 (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index).

PCR-RFLP technique and sequencing

TEX11 gene was amplified through PCR reaction. Forward and Reverse primers were designed by oligo7 version 7.56 (RYCHLIK, 2007) (Table1).

Table1. Forward and Reverse primers

Table 1.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

PCR reaction was carried out in total volume 25 µl containing 4 µl of DNA samples, 21 µl Master Mix (H2O, Buffer 10X, MgCl2 100 mM, dNTP-Mix 40 mM, Forward primer, Reverse primer and Taqpolymerase). Then, PCR was performed In thermo cycler system (Genetix Biotech, Australia). The PCR fragments were separated by 1% agarose gel electrophoresis and visualized by DNA power load. PCR products of 30 persons both from case and control were randomly selected for sanger sequencing. These individuals were from 4 Iranian ethnic: Turk, Fars, Lor and Kurd respectively (samples information Table3).

Table 2. Restriction enzyme information

NO	SNP	R.E name	R.E site	Cutting site number	company
1	rs775667438	BbvI	GCAGC	2	NEBcutter
2	rs746940663	BtsIMutI	C A G T G NN	1	NEBcutter
3	rs6525433	EarI	CTCTTC	1	NEBcutter
4	rs756553436	BceAI	ACGGC	1	NEBcutter
5	rs867296518	TSP45I	GTSAC	1	NEBcutter
6	rs977597709	SSPI	AATATT	1	NEBcutter
7	rs1056191384	BsmAI	GTCTC	2	NEBcutter

SNP genotyping was performed by RFLP technique and various restriction enzymes were chosen by NEB cutter website (http://nc2.neb.com/NEBcutter2/) for detecting each of those 7 SNPs (Enzymes information Table2). The PCR products were digested with restriction endonuclease enzymes (Thermo Fisher Scientific, United States) according to the manufacturers

protocol. The enzyme digestion mixture was carried out in total volume 15.5 µl containing 5 µl of PCR product, 9 µl double-distilled water (ddH_2O), 1 µl of 10X Buffer Tango and 0.5 µl Bsu36I enzyme (10 u/µl). The digested products were detected on 2% agarose gels electrophoresis with 50-1500 bp DNA ladder.

Table 3. Samples phenotype information

Sample number	Case/ control	Ethnicity	Age	Sperm/ ML	Volume	Total Sperm	morphology	mobility	PH	**viability**
1	control	fars	29	25	2	50	7	60	7.5	60
2	control	Fars	30	35	2	70	5	50	7.5	30
3	control	Fars	30	25	3	75	7	50	7.4	30
4	control	Fars	32	15	4	60	4	50	7.6	32
5	control	Fars	30	35	3	105	7	50	7.6	70
6	control	Turk	30	35	3	105	5	50	7.5	30
7	control	Turk	30	20	4	80	5	70	7.5	30
8	control	Turk	29	25	3	75	4	50	7.5	29
9	control	Turk	27	30	4	120	5	50	7.5	27
10	control	Turk	32	20	4	80	4	40	7.5	32
11	control	Lor	31	30	4	120	6	45	7.5	31
12	control	Lor	30	15	3	45	5	45	7.4	67
13	control	Kurd	30	35	2	70	5	50	7.6	30
14	control	Kurd	30	25	3	75	5	60	7.5	74
15	control	Kurd	30	25	4	100	6	50	7.5	75
16	CASE	Fars	30	0	3	0	0	0	7.5	0
17	CASE	Fars	31	0	4	0	0	0	7.5	0
18	CASE	Fars	30	0	3	0	0	0	7.5	0
19	CASE	Fars	31	0	4	0	0	0	7.5	0
20	CASE	Fars	31	0	3	0	0	0	7.4	0
21	CASE	Turk	28	0	3	0	0	0	7.4	0
22	CASE	Turk	30	0	3	0	0	0	7.5	0
23	CASE	Turk	30	0	4	0	0	0	7.5	0
24	CASE	Turk	30	0	3	0	0	0	7.4	0
25	CASE	Turk	30	0	3	0	0	0	7.5	0
26	CASE	Lor	28	0	3	0	0	0	7.5	0
27	CASE	Lor	30	0	3	0	0	0	7.5	0
28	CASE	Kurd	27	0	2	0	0	0	7.5	0
29	CASE	Kurd	34	0	3	0	0	0	7.5	0
30	CASE	Kurd	30	0	3	0	0	0	7.5	0

Data analyses

For association study, we used $\chi 2$ test as performed in SPSS Ver.21 (2000) software. To investigate the effect of ethnic population background on the phenotypes, we used clustering multivariate statistical methods, as X-chromosome SNPs are in hemizygous condition, and the

genotypes contain only one copy of the allele. In this situation, we cannot perform Hardy-Weinberg equilibrium (HE). The haplotype groups and haplotype diversity was determined by networking as performed in in Split trees 4, while ethnic group admixture was checked by POPART (Population Analysis with Reticulate Trees) program (http://popart.otago.ac.nz).

The sequence analysis (nucleotide diversity, number of segregating sites, number of parsimony-informative sites and Tajima's D statistic) among the studied subjects was done using MEGA7 (KUMAR *et al.,* 2016). Distribution versus close relationship between nucleotides was checked by detrended correspondence analysis (DCA) in PAST software.

The time of divergence for the studied ethnic groups was determined by BEAST v1.6.1 (DRUMMOND *et al.,* 2010a, b). BEAUti (Bayesian Evolutionary Analysis Utility version) v1.6.1 (DRUMMOND *et al.,* 2012a, b) was utilized to generate initial xml files for BEAST. A Yule process of speciation ('a pure birth' process) was used as a tree prior for all the tree model analyses.For the MCMC posterior analyses, the length of chain was 10 000 000. After 100 trees burn-in processing, 10 000 trees were used for the analyses. The BEAUti xml file was run in the BEAST v1.6.1 program and the maximum clade credibility (MCC) chain generations were repeated five times for each molecular clock model with independent runs to ensure suitable convergence and adequate mixing. The MCC tree was generated under the relaxed clock model (HKY substitution). We used a rate of evolution of the X-chromosome sequence (0.05). Tracer v1.5 software (RAMBAUT and DRUMMOND, 2007) was used for the output of the model parameters to examine the sampling and convergence results obtained from BEAST. Tree Annotator v1.6.1 software (DRUMMOND *et al.,* 2012a, b) was used to annotate the phylogenetic results generated by BEAST as a form of single 'target' tree. RASP (Reconstruction Ancestral State in Phylogenies) program were used (YU *et al.,* 2015), was used to identify the ancestral areas of ethnic group distribution, based on MCMC and Bayesian approach.

RESULTS

PCR-RFLP

All 160 samples were examined by RFLP technique for those 7 SNPs (Table4) and the χ^2 test did not show their associated with azoospermia male infertility (P = 0.9). Therefore, the studied TEXT-11 SNPs play no role in azospermia in the studied Iranian samples**.**

Table 4. RFLP results showing allele composition in the studied TEXT-11 SNPs

SNP	Case(Allelic type)	Control(Allelic type)
rs775667438	T	T
rs746940663	C	C
rs6525433	T	T
rs756553436	A	A
rs867296518	A	A
rs977597709	T	T
rs1056191384	A	A

Sequence analysis and population genetics

Sequencing of the randomly selected individuals within the studied samples, after alignment and curation, produced 87 bp DNA segment for further analysis (Fig. 1).

Figure 1. DNA segment with 87 bp length obtained after curation in the studied subjects.

The studied subjects differed in their nucleotide contents. The preliminary statistics of these sequences are: The nucleotide diversity = pi = 0.39, the number of segregating sites = 87, and the number of parsimony-informative sites = 66.

Tajima's D statistics produced D = 2.15, p (D >= 2.15174) = 0.01. Tajima's D is a population genetic test statistic that is the difference between the mean number of pairwise differences of DNAs and the number of segregating sites. If it becomes equal to zero (0), it shows that the population is neutrally evolving population with constant size. However, D values with significantly higher or lower than zero, indicates that the population evolve under a non-random process, including directional selection or balancing selection, demographic expansion or contraction, genetic hitchhiking, or introgression.

Tajima's D statistics >0, if it is statistically significant, reveal that rare alleles are lacking in the populations and probably the studied population is under the effect of balancing selection. This is the case with present study as we obtained D = 2.15, p (D >= 2.15174) = 0.01, which fits to the lack of rare alleles and show the potential effect of balancing selection in the studied subjects.

For population genetic investigation, molecular markers/or nucleotides used should be independent and represent a diverse poison in the genome. Therefore, DCA (Dentrented Correspondence Analysis) plot was obtained for the studied nucleotides (Fig. 2). The result showed that these nucleotides are well separated and scattered in the genome and therefore, are suitable molecular markers for differentiating the studied subjects and ethnic groups.

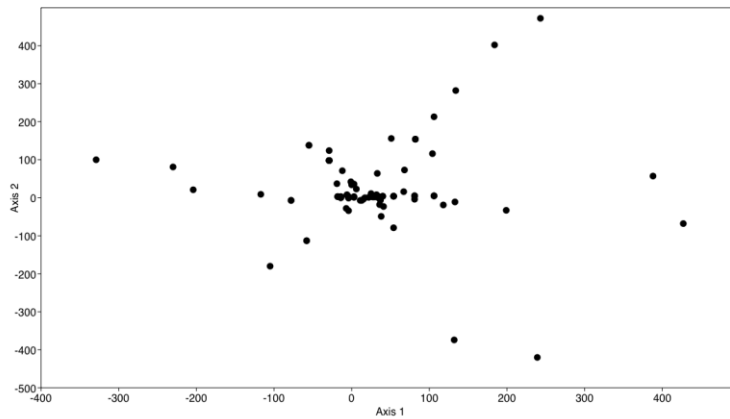

Figure 2. DCA plot of the studied nucleotides in azoospermia subjects.

Genetic diversity and haplotype groups

In the first attempt we investigated haplotype groups for the studied SNPs within all the studied samples (both control and patients) (Fig.3). In total 5 haplotype groups were identified. Each of these groups also had some degree of within group genetic variability. For example, in haplotype group 1, the subjects n4 and n9 had a higher degree of genetic similarity compared to the other subjects of the same group.

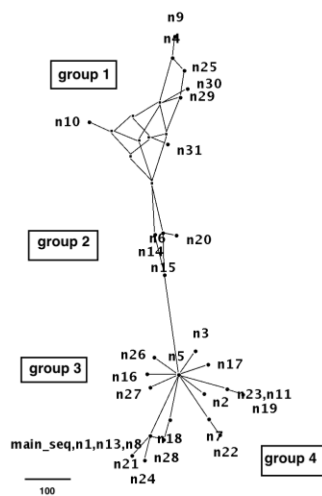

Figure 3. Haplotype groups identified in the studied subjects.

We also determined genetic distance based on Kimura 2-parameters among the studied subjects (Fig.4). The value obtained varied from 0.0 to 0.79, with average value of about 0. 2. This in general shows a very low degree of nucleotide replacement which is in agreement with low substitution rate reported for X-chromosome (about 0.05).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1 n31	–																					
2 n10	0.03	–																				
3 n29	0.02	0.04	–																			
4 n25	0.11	0.13	0.08	–																		
5 n20	0.41	0.39	0.38	0.48	–																	
6 n30	0.28	0.28	0.26	0.29	0.72	–																
7 n9	0.28	0.25	0.25	0.24	0.79	0.36	–															
8 n4	0.13	0.13	0.11	0.09	0.51	0.25	0.14	–														
9 n22	0.16	0.17	0.15	0.17	0.54	0.36	0.32	0.19	–													
10 n14	0.11	0.10	0.09	0.16	0.45	0.34	0.37	0.20	0.22	–												
11 n27	0.15	0.15	0.13	0.19	0.45	0.42	0.42	0.24	0.28	0.16	–											
12 n7	0.11	0.11	0.09	0.09	0.43	0.32	0.29	0.13	0.13	0.10	0.13	–										
13 n16	0.08	0.08	0.07	0.12	0.39	0.34	0.32	0.16	0.20	0.09	0.10	0.07	–									
14 n18	0.12	0.12	0.10	0.16	0.41	0.38	0.30	0.15	0.25	0.13	0.09	0.10	0.08	–								
15 n28	0.11	0.09	0.09	0.15	0.41	0.34	0.30	0.15	0.25	0.13	0.12	0.10	0.08	0.04	–							
16 n23	0.08	0.08	0.07	0.12	0.35	0.32	0.32	0.16	0.20	0.09	0.08	0.07	0.04	0.05	0.05	–						
17 n6	0.04	0.04	0.03	0.12	0.31	0.27	0.28	0.13	0.17	0.08	0.09	0.08	0.05	0.07	0.07	0.03	–					
18 n15	0.06	0.05	0.04	0.11	0.33	0.29	0.30	0.15	0.19	0.07	0.08	0.07	0.04	0.05	0.05	0.02	0.01	–				
19 n26	0.10	0.09	0.08	0.13	0.37	0.34	0.34	0.17	0.22	0.10	0.09	0.08	0.05	0.07	0.07	0.03	0.04	0.03	–			
20 n17	0.11	0.10	0.08	0.14	0.39	0.36	0.35	0.19	0.20	0.12	0.10	0.09	0.07	0.05	0.05	0.04	0.05	0.04	0.05	–		
21 n11	0.08	0.08	0.07	0.12	0.35	0.32	0.32	0.16	0.20	0.09	0.08	0.07	0.04	0.05	0.05	0.00	0.03	0.02	0.03	0.04	–	
22 n5	0.07	0.07	0.05	0.10	0.33	0.30	0.30	0.15	0.19	0.08	0.07	0.05	0.03	0.04	0.04	0.01	0.02	0.01	0.02	0.03	0.01	
23 n3	0.08	0.08	0.07	0.12	0.35	0.32	0.32	0.16	0.20	0.09	0.08	0.07	0.04	0.05	0.05	0.02	0.03	0.02	0.03	0.04	0.02	0.01
24 n19	0.10	0.09	0.08	0.13	0.37	0.31	0.34	0.18	0.20	0.10	0.09	0.08	0.05	0.07	0.07	0.01	0.04	0.03	0.04	0.05	0.01	0.02
25 n2	0.08	0.08	0.07	0.12	0.35	0.32	0.32	0.16	0.20	0.09	0.08	0.07	0.04	0.05	0.05	0.02	0.03	0.02	0.03	0.04	0.02	0.01
26 n21	0.08	0.07	0.07	0.12	0.37	0.30	0.30	0.15	0.22	0.10	0.09	0.08	0.04	0.07	0.04	0.03	0.04	0.03	0.04	0.05	0.03	0.02
27 main seq	0.07	0.05	0.05	0.10	0.35	0.29	0.28	0.13	0.20	0.09	0.08	0.07	0.04	0.05	0.03	0.02	0.03	0.02	0.03	0.04	0.02	0.01
28 n1	0.07	0.05	0.05	0.10	0.35	0.29	0.28	0.13	0.20	0.09	0.08	0.07	0.04	0.05	0.03	0.02	0.03	0.02	0.03	0.04	0.02	0.01
29 n13	0.07	0.05	0.05	0.10	0.35	0.29	0.28	0.13	0.20	0.09	0.08	0.07	0.04	0.05	0.03	0.02	0.03	0.02	0.03	0.04	0.02	0.01
30 n8	0.07	0.05	0.05	0.10	0.35	0.29	0.28	0.13	0.20	0.09	0.08	0.07	0.04	0.05	0.03	0.02	0.03	0.02	0.03	0.04	0.02	0.01
31 n24	0.08	0.07	0.07	0.12	0.37	0.30	0.30	0.15	0.22	0.10	0.09	0.08	0.05	0.07	0.04	0.03	0.04	0.03	0.04	0.05	0.03	0.02

Kimura 2-parameter distance matrix (continued)

	23	24	25	26	27	28	29	30	31
23 n3	–								
24 n19	0.03	–							
25 n2	0.02	0.03	–						
26 n21	0.03	0.04	0.03	–					
27 main seq	0.02	0.03	0.02	0.01	–				
28 n1	0.02	0.03	0.02	0.01	0.00.	–			
29 n13	0.02	0.03	0.02	0.01	0.00	0.00	–		
30 n8	0.02	0.03	0.02	0.01	0.00	0.00	0.00	–	
31 n24	0.03	0.04	0.03	0.02	0.01	0.01	0.01	0.01	–

Figure 4. Kimura 2-parameter distance matrix in the studied subjects.

Association study based on sequence data

Though we had obtained no association between the studied SNPs and azospermia by RFLP analysis. There are possibilities that neighboring nucleotides of the studied SNPs play some sort of role in male infertility. Therefore, we performed multivariate analyses of the DNA sequences to investigate their association with azoospermia. Both clustering and ordination methods, revealed that the studied SNPs could not differentiate the case and control subjects from each other which indicates these SNPSs and their neighboring nucleotides are not associated with male infertility. However, the UPGMA dendrogram of the sequences in SNP 1, in Fars ethinc population (Fig.5), as well as in Kurd ethnic group separated the case and control subjects in two distinct clusters. This may indicate that the neighboring nucleotides of this SNP, may play role in male infertility. Based on similar investigation, potential association between the studied SNPs and their neighboring sequences are provided in Table 5. All these results suggest that non-coding sequence of the studied SNPs may play role in causing male infertility in some of the ethnic populations.

Table 5. Details of ethnic populations in which SNP sequences were potentially associated with the male infertility

Number	SNP	Fars	Lor	Kurd	Turk
1	RS652533	-	+	-	-
2	RS1056191384	-	+	+	-
3	RS746940663	+	-	-	-
4	RS977597709	+	+	-	-
5	RS775667438	+	-	+	+
6	RS867296518	+	-	-	+
7	RS756553436	-	-	-	-

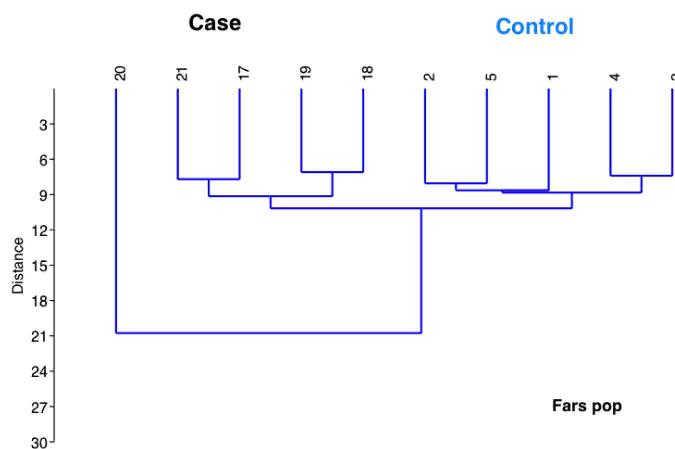

Figure 5. UPGMA dendrogram of the sequences in SNP 1, and in Fars ethinc population.

AMOVA performed based on combined sequences of all 7 SNPs, did not produced significant difference among the ethnic populations (Phipt = 0.2, P = 0.26). It reveals that no genetic differentiation occurred among ethnic populations, possibly due to gene flow and migration. TCS networking of the sequence data (Fig. 6) also shows high degree of admixture among the studied ethnic populations.

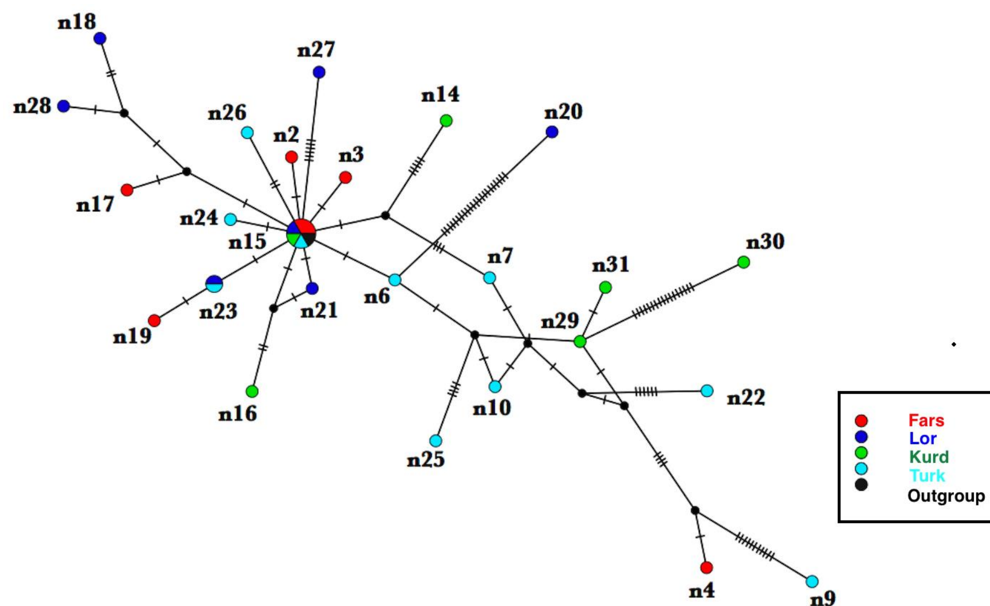

Figure 6. TCS networking of the studied subjects based on their ethnic groups, showing admixture of the subjects due to frequent gene flow.c.

Divergence time and ancestral distribution area

Divergence time obtained for the studied ethnic populations is presented in the chronological tree of BEAST (Fig. 7). The same tree was used for RASP (Reconstruction of ancestral area) analysis (Fig. 7). Divergence time obtained for each SNP differed to some degree from the others, as it occurs usually in X-Chromosome independent SNPs. However, in general and based on all SNP sequences studied, the time obtained for SNPs and ethnic populations varied from 1000- 30000 years ago, with 10000 years ago as active differentiation time.

Although the chronological tree obtained based on Bayesian algorithm of BEAST, shows that members of different ethnic populations are somewhat intermixed in clades (A, B, C, and D in Fig.7), the Lord population samples (C in Fig. 7) studied become mostly differentiated by 10000-15000 years ago. Similarly, the studied samples in Fars population, became differentiated around 8000 years ago (A in Fig. 7) The other two ethnic populations viz. Kurd and Turk (B and D, in Fig. 7), have much younger divergence time, around 1000-3000 years ago. The present investigation indicates that gene flow occurred among all these populations, and probably the two ethnic groups of Lor and Kurd that are located close to each other in West of Iran, had no limitation for migration and gene exchange. It also occurred towards south of Iran towards Fars population, and also towards North-West (Turk population). The reverse rote of migration may have occurred simultaneously.

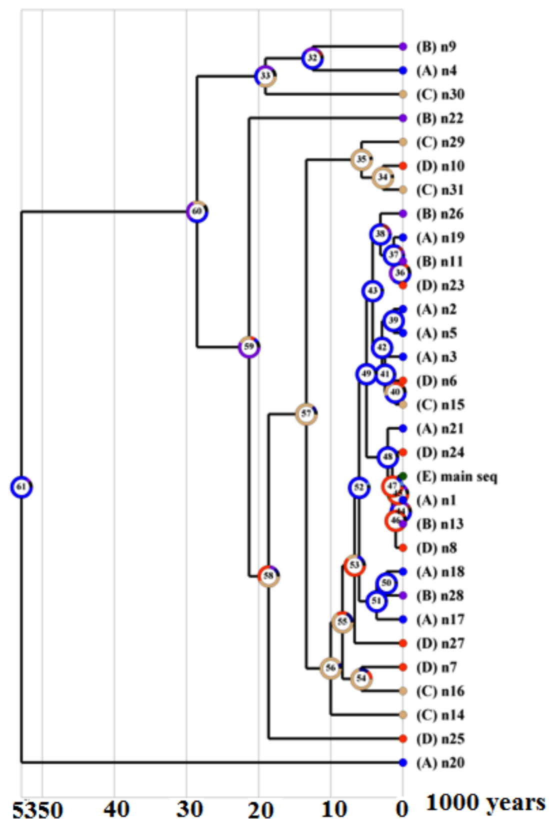

Figure 7. RASP tree of the studied subjects

DISCUSSION

Testis-expressed gene 11 (TEX11) is an integral component of the synaptonemal complex and ensures the progress of synapsis (YANG *et al.,* 2015). For example, TEX11 knockout mice fail to complete synapsis and crossover formation, which might result in male infertility, and loss of function of TEX15 in male mice severely impairs the meiotic recombination. The present study showed probable association of sequence variability in non-coding sequence of the studied SNPs in X-chromosme with male infertility within the studied ethnic groups.

In a similar investigation, YANG *et al*. (2015) suggested that the SNP polymorphisms TEX11, rs6525433, MLH3 rs175080, rs6525433–rs4844247, and rs1800734–rs175080 are associated with idiopathic male infertility presenting as azoospermia or oligozoospermia. However, many different genetic variants and other factors are also likely to influence male infertility. Further studies with larger numbers of subjects and different ethnic populations are needed to confirm these findings.

We observed almost a moderate level of genetic variability in sequences adjacent to X-chromosome SNPs. Moreover, grouping of the patients versus controls was associated with these sequences variability. 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%. Among the studied loci, pi value varied by over one order of magnitude. 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*. (2015b), 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.

We obtained some-what different divergence time tree for each of the studied SNPs. X chromosome contains many independent loci, each with its own phylogenetic tree. The phylogenetic trees for two loci can be very different, both in shape and in depth. Therefore, to have better insight on time of divergence and to unravel the complexities of the history of populations, it is preferred to use many loci. For this reason, we constructed a divergence time tree based on all obtained sequences in the 7 SNPs. The result is in general agreement with the known human population divergence time for example, GARRIGAN *et al*. (2007) estimated parameters of a general isolation-with-migration model by using sequence data from mitochondrial DNA (mtDNA), the Y chromosome, and two loci on the X chromosome in 10 humane populations. Results from comparisons between sub-Saharan African and Eurasian populations estimated that 1500 individuals founded the ancestral Eurasian population 40 thousand years ago (KYA). Analyses of sub-Saharan African populations suggest that they began diverging from one another upward of 50 KYA. Divergence times reported between non-African populations is about 40 KYA. The minimum divergence time between non-African populations is from the Dutch–Italian comparison, which yields a marginal posterior distribution with a mode at 7 KYA. The deepest non-African divergence times are the Dutch–Mongolian comparison (25 KYA) and the Baining– Papuan comparison (24 KYA). Finally, European and Asian populations were the first non-African populations to diversify, beginning 25 KYA. This divergence time is closely followed by divergence between the two Oceanian populations 24 KYA. Within Asia and Europe, divergence times are estimated to be more recent, ranging from 7 to 13 KYA and were accompanied by high levels of population growth (GARRIGAN *et al.,* 2007)

The use of the X chromosome in population genetics is still in its infancy (HEYER *et al.,* 2012; VERDU *et al.,* 2013), 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 non-coding sequences in genetic variability studies of ethnic groups and also portrait their potential association with idiopathic male infertility

Received, July 14th, 2022.
Accepted February 21st, 2023.

REFERENCE

COOPER, T.G., E., NOONAN, S., VON ECKARDSTEIN, J., AUGER, H.W.G., BAKER, H.M., BEHRE, T.B, HAUGEN, T.H., KRUGER, C.H., WANG, M.T., MBIZVO, K.M., VOGELSONG (2010): World Health Organization reference values for human semen characteristics. Human reproduction update, *16*(3): 231-245.

DRUMMOND, A.J., A., RAMBAUT, W., XIE (2011): BEAUti version 1.6.1. http://beast.bio.ed.ac.uk. Accessed 20 Apr.

DRUMMOND, A.J., A., RAMBAUT, M.A., SUCHARD (2003): BEAST version 1.6.1. http://beast.bio.ed.ac.uk. Accessed 13 March

EMISON, E.S, A.S., MCCALLION, C.S., KASHUK, R.T., BUSH, E., GRICE, S., LIN, M.E., PORTNOY, D.J., CUTLER, E.D., GREEN, A.A., CHAKRAVARTI (2005): common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. Nature, *434*: 857-863.

GARRIGAN, D., S.B., KINGAN, M.M., PILKINGTON, J.A., WILDER, M.P., COX, H.S., SOODYALL, B., STRASSMANN, G., DESTRO-BISOL, P., DE KNIJFF, A., NOVELLETTO, J., FRIEDLAENDER, M.F., HAMMER (2007): Inferring Human Population Sizes, Divergence Times and Rates of Gene Flow from Mitochondrial, X and Y Chromosome Resequencing Data. Genetics, *177*: 2195–2207.

GHIEH, F., V., MITCHELL, B., MANDON-PEPIN, F., VIALARD (2019): Genetic defects in human azoospermia. Basic Clin Androl., *29*: 4.

HARRIS, E.E, J.X., HEY (1999): chromosome evidence for ancient human histories. Proc. Natl. Acad. Sci. USA, *96*: 3320–3324.

HULL, M.G., C.M., GLAZENER, N.J., KELLY, D.I., CONWAY, P.A., FOSTER, R.A., HINTON, C., COULSON, P.A., LAMBERT, E.M., WATT, K.M., DESAI (1985): Population study of causes, treatment, and outcome of infertility. Br. Med. J., *291*:1693.

JARVI, K., K., LO, A., FISCHER, J., GRANTMYRE, A., ZINI, V., CHOW, V., MAK (2010): CUA Guideline: The workup of azoospermic males. CUAJ, *4*: 163–167

KAESSMANN, H., F., HEISSIG, A., VON HAESELER, S., PÄÄBO (1998): DNA sequence variation in a non-coding region of low recombination on the human X chromosome. Nature Genet., *22*: 78–81.

KUMAR, S., G., STECHER, K., TAMURA (2016): MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol., *33*: 1870–1874.

MATZUK, M.M., D.J., LAMB (2008): The biology of infertility: research advances and clinical challenges. Nat. Med., *14*:1197–1213.

NACHMAN, M.W., V.L., BAUER, S.L., CROWEL, C.F., AQUADRO (1999): DNA variability and recombination rates at X-linked loci in humans. Genetics, *150*: 1133-1141.

PONTING, C., G., LUNTER (2006): Signatures of adaptive evolution within human noncoding sequence. Hum Mol Genet., *15*: 170–175.

RAMBAUT, A., A., DRUMMOND (2007): Tracer, version.1.5. Molecular Evolution, Phylogenetics and Epidemiology Web. http://beast. bio. ed. ac. uk/. Accessed 2015

REIJO, R., T.Y., LEE, P., SALO, R., ALAGAPPAN, L.G., BROWN, M., ROSENBERG, S., ROZEN, T., JAFFE, D., STRAUS, O., HOVATTA, A., CHAPELLE, S.H., SILBER, D.C., PAGE (1995): Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA–binding protein gene. Nat Genet., *10*: 383–393.

SALAMONSEN, L.A., T., EDGELL, L.J.F., ROMBAUTS, M.D., ANDREW, N., STEPHENS, D.M., ROBERTSON, A., RAINCZUK, G., NIE, N.J., HANNAN (2013): Fertil Steril., *99*: 1086-1092.

SHEIDAI, M., M., HEYDARI, G., ATRI, N., KALHOR, F., KOOHDAR, Z., NOORMOHAMMADI (2021): 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.

TSAGAKIS, I., K., DOUKA, I., BIRDS, J.L., ASPDEN (2020): Long non-coding RNAs in development and disease: conservation to mechanisms. J. Pathol., *250*:480–495.

YANG, F., K., GELL, G.W., VAN DER HEIJDEN, S., ECKARDT, N.A., LEU, D.C., PAGE (2008): Meiotic failure in male mice lacking an X-linked factor. Genes & development, *22*(5):682-91.

YANG, F., SH., SILBER, N.A., LEU, R.D., OATES, J.D., MARSZALEK, H., SKALETSKY, L.G., BROWN, S., ROZEN, D.C., PAGE, P.J., WANG (2015) : TEX11 is mutated in infertile men with azoospermia and regulates genome-wide recombination rates in mouse. EMBO Mol. Med., *7*: 1198–1210.

YU, Y., A.J., HARRIS, C., BLAIR, X.J., HE (2015): RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. Mol. Phylogenet. Evol., *87*: 46–49.

ZHANG, S., Y., BIAN, L., LI, K., SUN, Z., WANG, Q., ZHAO, L., ZHA, J., CAI, Y., GAO, C., JI, C., LI (2015b): Population genetic study of 34 X-Chromosome markers in 5 main ethnic groups of China. Scientific Reports, *5*:17711.

ZHANG, X., M., DING, X., DING, T., LI, H., CHEN (2015a): Six polymorphisms in genes involved in DNA double-strand break repair and chromosome synapsis: association with male infertility. Syst. Biol. Reprod. Med., *61* (4): 187–193.

ZHENG, K., F., YANG, P.J., WANG (2010): Regulation of Male Fertility by X-Linked Genes. J. Androl., *31*:79-85.

TEXT-11 SNP-ovi I ULOGA SUSEDNIH SEKVENCI U IZAZIVANJU MUŠKE NEPLODNOSTI KOD NEKIH ETNIČKIH GRUPA IRANACA I NJIHOVA POTENCIJALNA ULOGA U PROCENI VREMENA DIVERGENCIJE

Masoud SHEIDAI[1], Sogol ALLAHVERDI[2], Narges ANVARI[2], Somayeh EYVAZI KHAMENE[2], Mozhgan GHANDEHARI[2], Marziyeh GHORESHI[2], Fatemeh JAMSHIDI[2], Gazal KHOSRAVANI[2], Naser KALHOR[3], Fahimeh KOOHDAR[1]

[1]Fakultet prirodnih nauka i biotehnologije, Univerzitet Shahid Beheshti, Teheran, Iran
[2]Ogranak Fakulteta za farmaceutske nauke, Islamski univerzitet Azad, Teheran, Iran
[3]Odeljenje za mezanhimalne matične ćelije, Akademski centar za obrazovanje, kulturu i istraživanje, Ogranak Kom, Iran

Izvod

X-hromozom i njegovi nezavisni lokusi su nedavno dobili značaj u genetskim istraživanjima koja se bave studijama kontrole bolesti, populacijskim genetskim analizama i filogenetskim istraživanjima. Neplodnost i nedostatak normalne reprodukcije su važna zdravstvena pitanja. Azoospermija je vrsta neplodnosti kod muškaraca čije seme ne sadrži spermu i može zahvatiti oko 1-20% muške populacije. Većina slučajeva spermatogenog neuspeha kod ljudi je idiopatska, a pretpostavlja se da su osnovni uzroci genetski. Analizirana je potencijalnu povezanost između sedam SNP-ova X-hromozoma i susednih sekvenci sa azoospermijom u četiri etničke populacije u Iranu. Takođe smo izvršili populacione genetičke analize da bismo po prvi put dobili podatke o genetskoj varijabilnosti, strukturi populacije, protoku gena i vremenu divergencije u ovim etničkim grupama. Kako su geni X-hromozoma/SNP-ovi u homozigotnom stanju, za asocijativno proučavanje koristili smo multivarijantne statističke metode, koje su otkrile potencijalnu ulogu varijabilnih mesta bliskih proučavanim SNP-ima u muškoj neplodnosti. Grupe haplotipova su identifikovane za svaki od sedam SNP-ova. Rezultati su pokazali protok gena među etničkim grupama usled migracije i međusobnog razmnožavanja, a Bajesova analiza podataka o sekvenci sugeriše da je vreme divergencije za etničke populacije između 2-30 KY.

Primljeno 1.VII.2022.
Odobreno 21. II 2023.