Reports of Biochemistry & Molecular Biology
Vol. 4, No. 2, Apr 2016

Short article

www.RBMB.net

Comparative *In silico* Study of Sex-Determining Region Y (SRY) Protein Sequences Involved in Sex-Determining

Masoume Vakili Azghandi[1], Mohammadreza Nasiri*[1], Ali Shamsa[2], Mohsen Jalali[2], Mohammad Mahdi Shariati[1]

Abstract

Background: The SRY gene (*SRY*) provides instructions for making a transcription factor called the sex-determining region Y protein. The sex-determining region Y protein causes a fetus to develop as a male. In this study, SRY of 15 spices included of human, chimpanzee, dog, pig, rat, cattle, buffalo, goat, sheep, horse, zebra, frog, urial, dolphin and killer whale were used for determine of bioinformatic differences.
Methods: Nucleotide sequences of *SRY* were retrieved from the NCBI databank. Bioinformatic analysis of *SRY* is done by CLC Main Workbench version 5.5 and ClustalW (http:/www.ebi.ac.uk/clustalw/) and MEGA6 softwares.
Results: The multiple sequence alignment results indicated that *SRY* protein sequences from *Orcinus* orca (killer whale) and *Tursiopsaduncus* (dolphin) have least genetic distance of 0.33 in these 15 species and are 99.67% identical at the amino acid level. *Homosapiens* and *Pantroglodytes* (chimpanzee) have the next lowest genetic distance of 1.35 and are 98.65% identical at the amino acid level.
Conclusion: These findings indicate that the *SRY* proteins are conserved in the 15 species, and their evolutionary relationships are similar.

Keywords: Bioinformatics analysis, *SRY* gene, Phylogeny

Introduction

In mammals, the Y chromosome-linked (sex determining factor, *SRY*) gene, is responsible for male sex determination. The SRY was discovered by analysis of the small fragments of the Y chromosome that had translocated to the X chromosome in the genomes of XX males and true hermaphrodites (1). *SRY* encodes a protein with a central HMG-box (High Mobility Group box) present in a wide variety of proteins that bind and bend DNA, suggesting that *SRY* functions as a transcription factor (2). The protein encoded by *SRY* showed sequence-specific DNA binding activity, which was absent or reduced in *SRY* from certain XY females with gonadal dysgenesis (3-5). The HMG-box shows sequence conservation with a heterogeneous group of nuclear proteins with diverse functions including transcriptional activation.

Although HMG-box sequences of *SRY* are reasonably conserved between the species studied to date, sequences outside the HMG-box display a notable lack of sequence conservation (6). The role of *SRY* as a transcription factor in sex determination in mammals remains elusive and other genes including SOX9, DMRT1, WNT1, AMH, SF1, DAX1, GATA4, LIM1, Fra1and aromatase also seem to be involved in the sex-determining pathway (7-9). The evolutionary analysis of the *SRY* coding region among primates and rodents suggests that this gene is rapidly evolving. In contrast, results from wallaby and domestic ruminants appear to indicate that sequence

1: Department of Animal Sciences, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran
2: Department of Urology, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran
***Corresponding author: Mohammadreza Nassiri; Tel: +989153114119; Fax: +985138803000; E-mail: nassiryr@um.ac.ir**
Received: Mar 15, 2015; Accepted: Jun 15, 2015

evolution of the *SRY* is less rapid (6). Zwingman *et al*. (10, 11) reported that the *SRY* is transcribed during mouse pre implantation development as early as the two-cell stage, while other sex determining regions are not, suggesting that mammalian sex determination starts prior to gonadal differentiation. The present study was conducted to determine the bioinformatic differences and relationship of *SRY* protein sequences from 15 different spices.

Materials and Methods

Obtaining protein sequences

The functional protein sequences of the SRY from the 15 species chosen were downloaded from the NCBI database (Table 1).

Multiple sequence alignment

These sequences were analyzed on CLC Main Workbench version 5.5 and ClustalW (http:/www.ebi.ac.uk/clustalw/) for the multiple sequences alignment (Fig. 1).

Table 1. Descriptions of SRY protein sequences from the 15 species analyzed in this study.

Seq. no.	Organism	Protein ID (NCBI Reference Sequence)	Locus	Definition	Version	Length (aa)
1	**Bostaurus (cattle)**	EU_581861.1	EU_581861	Sex determining region Y (SRY) gene	EU_581861.1 GI:190333256	5021
2	**Bubalusbubalis (water buffalo)**	DQ_119747.1	DQ_119747	Sex determining factor (SRY) gene	DQ_119747.1 GI:71082649	281
3	**Canis lupus familiaris (dog)**	AF_107021.1	AF_107021	Sex determining region Y protein (SRY) gene	AF_107021.1 GI:5114117	526
4	***Capra hircus* (goat)**	JN_561348.1	JN_561348	Sex-determining region of Y protein gene	JN_561348.1 GI:363990272	366
5	**Equuscaballus (horse)**	Z_26908.1	Z-26908	SRY gene	Z_26908.1 GI:407796	56
6	**Equusgrevyi (Grevy's zebra)**	EU_240941.1	EU_240941	Sex-determining region Y protein (SRY) gene	EU_240941.1 GI:160221277	330
7	***Homo sapiens* (human)**	JQ_811934.1	JQ_811934	SRY (SRY) gene	JQ_811934.1 GI:383087990	274
8	**Musmusculus (mouse)**	U_70654.1	U_70654	Sex determining protein (SRY) gene	U_70654.1 GI:2623372	532
9	***Orcinus orca* (killer whale)**	AB108526.2 GI:134287141	AB108526.2 GI:134287141	SRY gene for testis determining factor	AB108526.2 GI:134287141	641
10	**Ovisaries (sheep)**	Z_30265.1	Z_30265	Sex determining region of Y chromosome	Z_30265.1 GI:607141	241
11	**Ovisvignei (Urial)**	JN_992678.1	JN_992678	Sex determining region Y(SRY) gene	JN992678.1 GI:361584461	241
12	**Pan troglodytes (chimpanzee)**	DQ_977342.1	DQ_977342	determining region Y(SRY) gene	DQ_977342.1 GI:124111148	381
13	**Rattusnorvegicus (Norway rat)**	FJ_168067.1	FJ_168067	SRY (SRY) gene	FJ_168067.1 GI:208463432	707
14	**Susscrofa (pig)**	GU_143246.1	GU_143246	transcription factor Sry1 (SRY1) gene	GU_143246.1 GI:313755232	458
15	***Tursiopsaduncus* (Indo-pacific bottlenose dolphin)**	AB_275396.1	AB_275396	Sex determining region Y (SRY) gene	AB_275396.1 GI:134122709	641

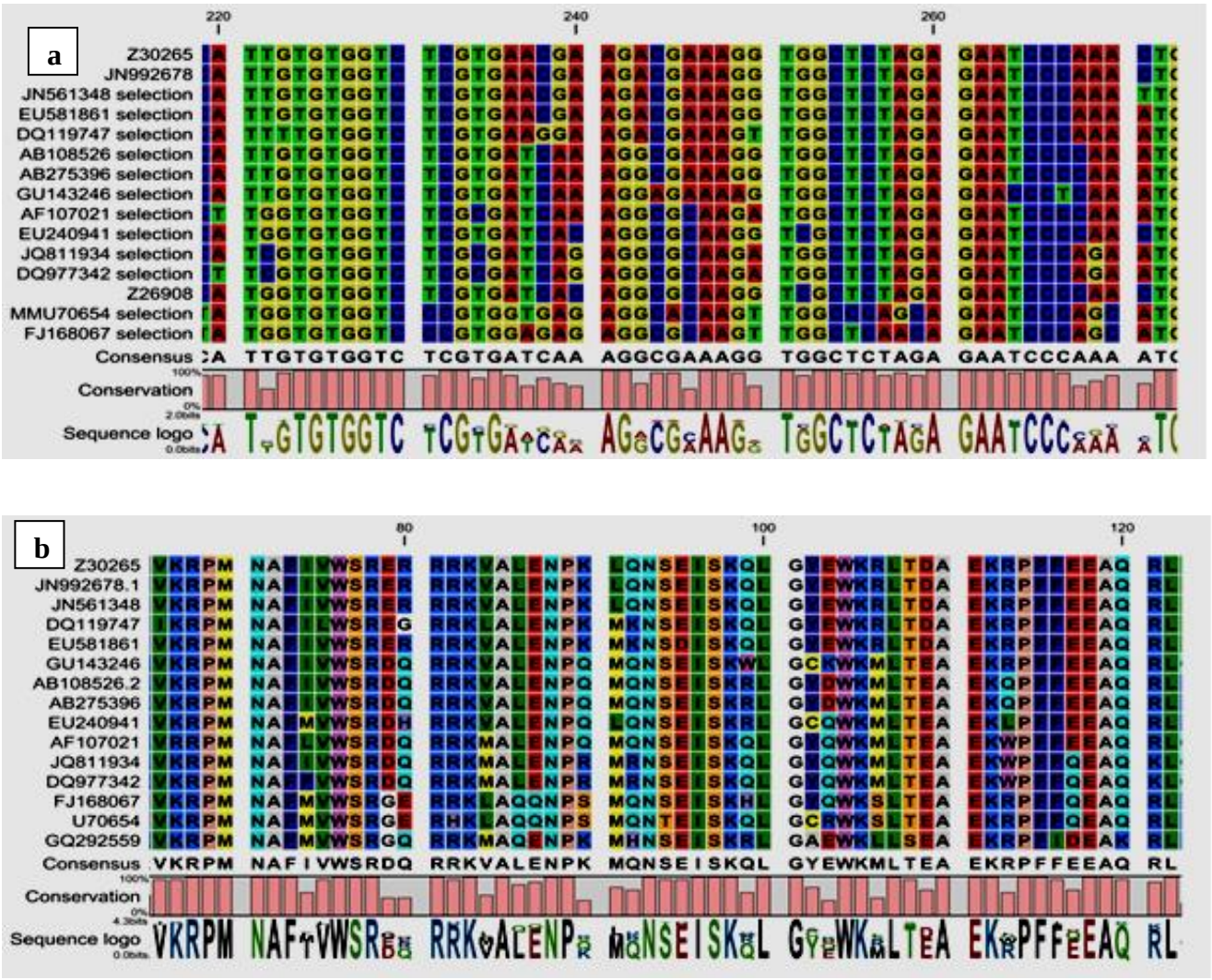

Fig. 1. Multiple sequence alignment result (a: DNA, b: Protein) by by Geneious 7.1.2 and CLC Main Workbench version 5.5. Green boxes represent 100% identity between species. Green-brown regions represent 30-99% identity, and red regions represent 0-29% identity.

Construction of phylogenetic trees

The phylogenetic trees were first constructed using the neighbor joining method from the MEGA6 package. Confidence on each node was assessed by 1000 bootstrap replications (Fig. 2). Also the maximum likelihood method from a MEGA6 package was used to construct a phylogenetic tree and 1000 replicates were used for bootstrap statistical test (Fig. 3).

Pairwise distances

To measure genetic distances between sequences, a pairwise distances method from the CLC Main Workbench version 5.5 package was used. SRY proteins and nucleotide in human, chimpanzee, dog, pig, rat, cattle, buffalo, goat, sheep, horse, zebra, frog, urial, dolphinand killer whale (showed by green in Fig. 1).

The pairwise distances method of phylogenetic analysis relies on a measure of genetic distance between the sequences being classified. This analysis shows the differences and percent identity of each sequence pair in the current alignment.

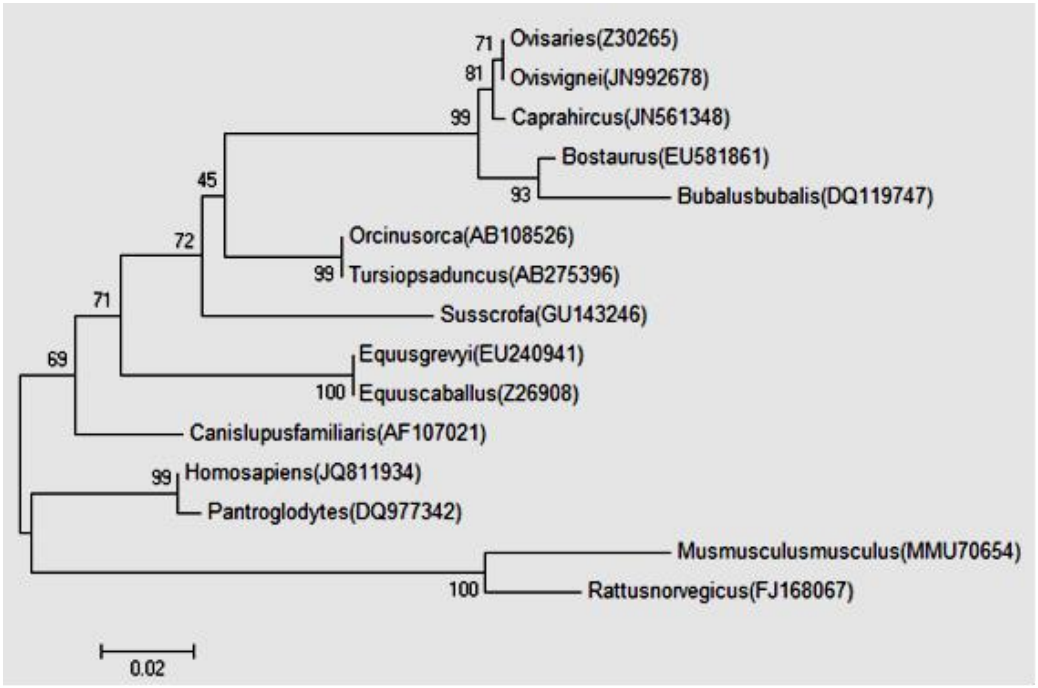

Fig. 2. Molecular phylogenetic analysis by the neighbor-joining method. The optimal phylogenetic tree with the sum of branch length 0.49450736 is shown.

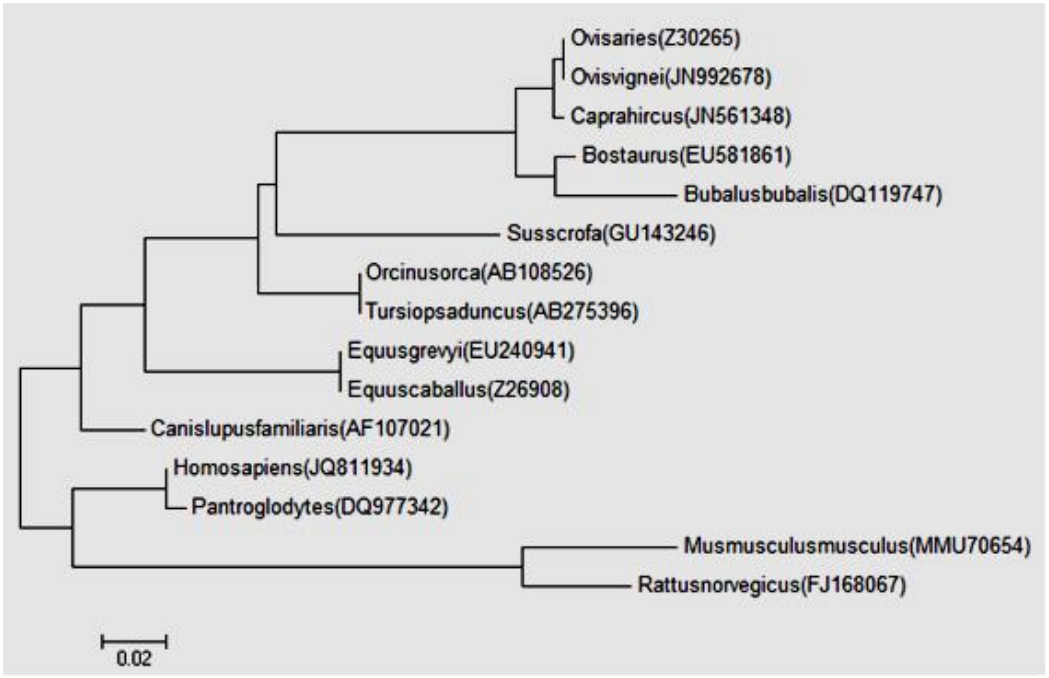

Fig. 3. Molecular phylogenetic analysis by the maximum likelihood method. The evolutionary history was inferred using the maximum likelihood method based on the JTT matrix-based model with 1000 replicates of the bootstrap test.

Results and Discussion

Sequence comparison between the *SRY* protein sequences (Fig. 4) indicated that the *SRY* protein Sequences from *Orcinus orca* (killer whale) and *Tursiopsaduncus* (dolphin) have least genetic distance of 0.33 in these 15 species and are 99.67% identical at the amino acid level. Sex-determining region Y protein sequences from *Homo sapiens* and Pan t*roglodytes* (chimpanzee) have a genetic distance of 1.35 and are 98.65% identical at the amino acid level. The maximum genetic distance of 89.98 occurred between *Equuscabal* and *Musmusculus* (mouse). These proteins were 10.02% similar at the amino acid level (Fig. 5).

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Ovis aries (Z30265)	1		1	36	44	51	198	198	166	220	209	302	298	547	1218	498
Ovis vignei (JN992678)	2	99.85		36	44	51	198	198	165	220	209	301	297	547	1218	498
Capra hircus (JN561348)	3	94.76	94.76		60	67	209	209	181	233	224	311	309	548	1221	499
Bos taurus (EU581861)	4	93.60	93.60	91.27		21	195	194	171	222	215	298	296	552	1217	492
Bubalus bubalis (DQ119747)	5	92.58	92.58	90.25	96.94		199	198	175	227	221	302	300	557	1219	494
Orcinus orca (AB108526)	6	71.47	71.47	69.88	71.90	71.33		2	199	200	240	208	208	457	1145	416
Tursiops aduncus (AB275396)	7	71.47	71.47	69.88	72.05	71.47	99.67		197	200	239	207	207	457	1145	416
Sus scrofa (GU143246)	8	76.08	76.22	73.92	75.36	74.78	71.20	71.49		232	209	306	302	541	1228	507
Canis lupus familiaris (AF107021)	9	68.66	68.66	66.81	68.38	67.66	69.83	69.83	66.81		191	245	244	511	1172	445
Equus grevyi (EU240941)	10	70.60	70.60	68.50	69.76	68.92	65.96	66.10	70.48	72.91		278	279	540	1207	484
Homo sapiens (JQ811934)	11	57.52	57.67	56.26	58.09	57.52	66.77	66.93	56.96	63.70	60.73		8	446	1123	396
Pan troglodytes (DQ977342)	12	58.09	58.23	56.54	58.37	57.81	66.77	66.93	57.52	63.85	60.59	98.65		447	1124	397
Equus caballus (Z26908)	13	20.38	20.38	20.23	19.65	18.92	24.34	24.34	20.91	22.58	23.40	24.53	24.37		1122	379
Mus musculus musculus (MMU70654)	14	18.96	18.96	18.76	19.03	18.90	19.42	19.42	18.30	20.16	19.75	20.30	20.23	10.02		804
Rattus norvegicus (FJ168067)	15	35.58	35.58	35.45	36.35	36.09	39.53	39.53	34.41	39.46	37.31	41.42	41.27	25.10	35.53	

Fig. 4. Estimates of evolutionary divergence between SRY protein sequences. Pairwise distances methods of phylogenetic analyses were conducted using the Poisson correction model. The analysis involved 15 amino acid sequences. All positions containing gaps and missing data were eliminated. This analysis shows the divergence of each sequence pair in the current alignment.

Phylogenetic tree

This analysis shows evolutionary relationships predicted from the multiple sequence alignment. The length of each pair of branches represents the distance between sequence pairs. The phylogenic trees (Fig. 3 and 4) classified the species into two groups. Group 1 contains of four species (*Homo sapiens, Pan troglodytes, Rattusnorvegicus, and Musmusculus*) with the lowest genetic distances, Group 2 contains11 species (*Canis lupus, Tursiopsaduncus, Susscrofa, Ovisvignei ,Equusgrevyi, Orcinus orca , Bostaurus, Bubalusbubalis, Capra hircus, Ovisaries, and Equuscaballus*).

The phylogenetic tree constructed by MEGA 6 shows that the evolutionary relationships between the *SRY* protein sequences from the 15 species analyzed are similar to their species evolutionary relationships**.**

Molecular Phylogenetic analysis by Maximum Likelihood method

The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model (12). The tree with the highest log likelihood (-755.3722) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. All positions containing gaps and missing data were eliminated. There were a total of 165 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (13).

Molecular Phylogenetic analysis by Neighbor-Joining method

The evolutionary history was inferred using the Neighbor-Joining method (14). The optimal tree with the sum of branch length 0.49450736 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (15). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method (14) and are in the units of the number of base substitutions per site. All positions containing gaps and missing data were eliminated. There were a total of 165 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (15).

The SRY proteins are conserved in the 15 species analyzed in our study, and their evolutionary relationships are similar to the species evolutionary relationships.

Acknowledgment

The authors gratefully acknowledge the contribution of Montaserieh Infertility Research & Treatment Center of Mashhad for assisting with the conduct of the study. This work was financially supported by Ferdowsi University of Mashhad.

References

1. Berta P, J.R. H, Sinclair AH, Taylor A, Griffiths BL, Goodfellow PN, et al. Genetic evidence equating SRY and the testis-determining factor. Nature. 1990;348(6300):448-50.
2. Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. Evolution. 1985;39(4):783-91.
3. Harley VR, Goodfellow PN. The biochemical role of SRY in sex determination. Molecular Reproduction and Development. 1994;39(2):184-93.
4. Harley VR, Jackson DI, Hextall PJ, Hawkins JR, Berkovitz GD, Sockanathan S, et al. DNA binding activity of recombinant SRY from normal males and XY females. Science. 1992;255(5043):453-6.
5. Koopman P. Sry and Sox9: mammalian testis-determining genes. Cellular and Molecular Life Sciences. 1999;55(6):839-56.
6. Koopman P, Munsterberg A, Capel B, Vivian N, Lovell-Badge R. Expression of a candidate sex-determining gene during mouse testis differentiation. Nature. 1990;348(6300):450-2.
7. Nasrin N, Buggs C, Kong XF, Camazza J, Goebl M, Alexander-Bridges M. DNA-binding properties of the product of the testis-determining gene and a related protein. Nature. 1991;354(6351):317-20.
8. Poulat F, Girard F, Chevron MP, Gozé C, Rebillard X, Calas B, et al. Nuclear localization of the testis determining gene product SRY. The Journal of Cell Biology. 1995;128(5):737-48.
9. Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. molecular Biology and Evolution. 1987;4(4):406-25.
10. Sinclair AH, P. B, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, et al. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature. 1990;346(6281):240-4.
11. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution. 1993;103:512-26.
12. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. molecular Biology and Evolution.30(12):2725-9.
13. Töhönen V, Frygelius J, Mohammadieh M, Kvist U, Pelliniemi LJ, .O'Brien K, et al. Normal sexual development and fertility in testatin knockout mice. Molecular and Cellular Biology. 2005;25(12):4892-902.
14. Whitfield LS, Lovell-Badge R, Goodfellow PN. Rapid sequence evolution of the mammalian sex-determining gene SRY. Nature. 1993;364(6439):713-5.
15. Zwingman T, Erickson RP, Boyer T, Ao A. Transcription of the sex-determining region genes Sry and Zfy in the mouse preimplantation embryo. Proceedings of the National Academy of Sciences. 1993;90(3):814-7.