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

Masoume Vakili Azghandi¹, Mohammadreza Nasiri*¹, Ali Shamsa², Mohsen Jalali², Mohammad Mahdi Shariati¹

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 Tursiops aduncus (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, Fra1 and 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
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 species.

**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.

<table>
<thead>
<tr>
<th>Seq. no.</th>
<th>Organism</th>
<th>Protein ID (NCBI Reference Sequence)</th>
<th>Locus</th>
<th>Definition</th>
<th>Version</th>
<th>Length (aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bostaurus (cattle)</td>
<td>EU_581861.1</td>
<td>EU_581861</td>
<td>Sex determining region Y (SRY) gene</td>
<td>EU_581861.1 GI:19033256</td>
<td>5021</td>
</tr>
<tr>
<td>2</td>
<td>Bubalus bubalis (water buffalo)</td>
<td>DQ_119747.1</td>
<td>DQ_119747</td>
<td>Sex determining factor (SRY) gene</td>
<td>DQ_119747.1 GI:71082649</td>
<td>281</td>
</tr>
<tr>
<td>3</td>
<td>Canis lupus familiaris (dog)</td>
<td>AF_107021.1</td>
<td>AF_107021</td>
<td>Sex determining region Y protein (SRY) gene</td>
<td>AF_107021.1 GI:5114117</td>
<td>526</td>
</tr>
<tr>
<td>4</td>
<td>Capra hircus (goat)</td>
<td>JN_561348.1</td>
<td>JN_561348</td>
<td>Sex-determining region of Y protein gene</td>
<td>JN_561348.1 GI:363990272</td>
<td>366</td>
</tr>
<tr>
<td>5</td>
<td>Equus caballus (horse)</td>
<td>Z_26908.1</td>
<td>Z_26908</td>
<td>SRY gene</td>
<td>Z_26908.1 GI:407796</td>
<td>56</td>
</tr>
<tr>
<td>6</td>
<td>Equus grevyi (Grevy’s zebra)</td>
<td>EU_240941.1</td>
<td>EU_240941</td>
<td>Sex-determining region Y protein (SRY) gene</td>
<td>EU_240941.1 GI:160221277</td>
<td>330</td>
</tr>
<tr>
<td>7</td>
<td>Homo sapiens (human)</td>
<td>JQ_811934.1</td>
<td>JQ_811934</td>
<td>SRY (SRY) gene</td>
<td>JQ_811934.1 GI:383087990</td>
<td>274</td>
</tr>
<tr>
<td>8</td>
<td>Mus musculus (mouse)</td>
<td>U_70654.1</td>
<td>U_70654</td>
<td>Sex determining protein (SRY) gene</td>
<td>U_70654.1 GI:2623372</td>
<td>532</td>
</tr>
<tr>
<td>9</td>
<td>Orcinus Orca (killer whale)</td>
<td>AB108526.2 GI:134287141</td>
<td>AB108526.2</td>
<td>SRY gene for testis determining factor</td>
<td>AB108526.2 GI:134287141</td>
<td>641</td>
</tr>
<tr>
<td>10</td>
<td>Ovis aries (sheep)</td>
<td>Z_30265.1</td>
<td>Z_30265</td>
<td>Sex determining region of Y chromosome</td>
<td>Z_30265.1 GI:607141</td>
<td>241</td>
</tr>
<tr>
<td>11</td>
<td>Ovisignei (Urial)</td>
<td>JN_992678.1</td>
<td>JN_992678</td>
<td>Sex determining region Y(SRY) gene</td>
<td>JN992678.1 GI:361584461</td>
<td>241</td>
</tr>
<tr>
<td>12</td>
<td>Pan troglodytes (chimpanzee)</td>
<td>DQ_977342.1</td>
<td>DQ_977342</td>
<td>determining region Y(SRY) gene</td>
<td>DQ_977342.1 GI:124111148</td>
<td>381</td>
</tr>
<tr>
<td>13</td>
<td>Rattus norvegicus (Norway rat)</td>
<td>FJ_168067.1</td>
<td>FJ_168067</td>
<td>SRY (SRY) gene</td>
<td>FJ_168067.1 GI:208463432</td>
<td>707</td>
</tr>
<tr>
<td>14</td>
<td>Sus scrofa (pig)</td>
<td>GU_143246.1</td>
<td>GU_143246</td>
<td>transcription factor Sry1 (SRY1) gene</td>
<td>GU_143246.1 GI:313755232</td>
<td>458</td>
</tr>
<tr>
<td>15</td>
<td>Tursiops aduncus (Indo-pacific bottlenose dolphin)</td>
<td>AB_275396.1</td>
<td>AB_275396</td>
<td>Sex determining region Y (SRY) gene</td>
<td>AB_275396.1 GI:134122709</td>
<td>641</td>
</tr>
</tbody>
</table>
**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, dolphin, and killer whale (showed by green in Fig. 1) were used.

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. 1.** Multiple sequence alignment result (a: DNA, b: Protein) 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.
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 troglodytes (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 Equus caballus and Mus musculus (mouse). These proteins were 10.02% similar at the amino acid level (Fig. 5).
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 phylogenetic 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 contains 11 species (Canis lupus, Tursiops aduncus, Suscrofa, 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 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