Arch. Biol. Sci., Belgrade, 60 (4), 525-530, 2008 DOI:10.2298/ABS0804525K

PCR AMPLIFICATION AND SEQUENCE ANALYSIS OF THE RAT *SOX3* GENE

A. KRSTIĆ, MARIJA MOJSIN, NATAŠA KOVAČEVIĆ-GRUJIČIČ, and MILENA STEVANOVIĆ

Institute of Molecular Genetics and Genetic Engineering, 11010 Belgrade, Serbia

Abstract — The *Sox3* gene is considered to be one of the earliest neural markers in vertebrates, playing a role in specifying neuronal fate. Despite the completion of a rat genome sequencing project, only a partial sequence of the rat *Sox3* gene has been available in the public database. Using PCR, sequencing, and bioinformatics tools, in this study we have determined the complete coding sequence of the rat *Sox3* gene encoding 449 amino acids. Comparative analysis of rat and human SOX3 proteins revealed a high degree of conservation. Identification of the rat *Sox3* gene sequence would help in understanding the biological roles of this gene and provide insight into evolutionary relationships with vertebrate orthologs.

Key words: *Rattus norvegicus*, laboratory rat, *Sox3*, comparative proteomics

UDC 577.21:599.323.4

INTRODUCTION

Sox genes comprise a large family of *SRY*-related HMG-box genes involved in the decision of cell fates during development and implicated in the control of diverse developmental processes (Pevny and Lovell-Badge, 1997; Wegner, 1999). SOX proteins display properties of both classical transcriptional factors and architectural components of chromatin (Pevny and Lovell-Badge, 1997).

Based on HMG box homology and intron–exon structure, *SOX* genes are divided into 10 distinct groups designated A-J (Bowles et al., 2000). The B group of *SOX* genes is of particular interest since members of this group play major roles in neural development. The five intron-less group B *SOX* genes (*SOX1*, *SOX2*, *SOX3*, *SOX14*, and *SOX21*) participate in the earliest events of CNS differentiation in *Drosophila*, *Xenopus*, chick, mouse, ascidians, and hemichordates (Sasai, 2001). Based on sequence analysis and functional studies in vertebrates, the group B *SOX* genes can be further subdivided into sub-group B1 comprising activators (*SOX1*, *SOX2*, and *SOX3*) and sub-group B2 consisting of repressors (*SOX14* and *SOX21*) (Uchikawa et al., 1999).

Sox3/SOX3, an X-linked member of the family, is considered to be one of the earliest neural markers in vertebrates, playing a role in specifying neuronal fate (Brunelli et al., 2003). Studies performed on *Sox3* null mice (Weiss et al., 2003) showed that this gene is important for normal oocyte development and male testis differentiation and gametogenesis. Deletion of the *Sox3* gene in mice resulted in defects of pituitary function and of specific CNS midline structures (Rizzoti et al., 2004). Evidence for developmental importance of the *SOX3* gene also comes from mutational analysis in humans. Dysfunction of the SOX3 protein disturbs cellular processes required for cognitive and pituitary development, leading to mental retardation and growth hormone deficiency in humans (Laumonnier et al., 2002; Stankiewicz et al., 2005; Woods et al., 2005).

Additionally, links have been found between the *SOX3* gene and cancers (Dong et al., 2004). Ectopic *SOX3* expression induced oncogenic transformation of chicken embryonic fibroblasts (Xia et al., 2000). In addition to several known proto-oncogenes, *Sox3* was identified in genome-based analysis of retroviral insertion sites in T-cell lymphomas (Kim et al., 2003). Up-regulation of the *SOX1*, *2*, *3,* and *21* genes

were found in lung squamous cell carcinomas (Gure et al., 2000). Taken together, these data suggest that the *SOX3* gene might play a role in tumorigenesis.

From the time when the human *SOX3* gene was first cloned and characterized (Stevanovic et al., 1993), ortholog genes have been cloned from various vertebrate species. Despite the complete rat genome sequencing, only a partial sequence of the rat *Sox3* gene has been available in public databases. Accordingly, the aim of this study was to determine the complete coding sequence of the rat *Sox3* gene and perform comparative proteomic analysis. Identification of the rat *Sox3* gene sequence would help in understanding the biological roles of this gene and provide insight into evolutionary relationships with vertebrate orthologs.

MATERIALS AND METHODS

Identification of partial rat Sox3 gene sequence

Rat genome sequences homologous to the human *SOX3* gene (AL121875.10) were searched for with the BLAST program (NCBI).

Preparation of genomic DNA

Frozen rat liver tissue (50 mg) was homogenized in 1 ml of a solution containing 50 mM Tris HCl (pH 9.0), 100 mM EDTA (pH 8.0), 200 mM NaCl, 1% SDS, and 10 μg/ml proteinase K. The homogenate was incubated overnight at 45°C, then treated with RNase A (10 μg/μl) for 1 h at 37°C. This was followed by phenol-chloroform and chloroform extraction. After precipitation with ethanol, the pellet was rinsed in 70% ethanol, dried, and dissolved in TE.

PCR amplification of rat genomic DNA

PCR primers used to amplify the rat *Sox3* gene were designed to match regions conserved between available rat and human *Sox3* gene sequences and to surround the existing gap in the rat sequence (contig NW_001091925.1). The sequences of primers are as follows:

F27: 5' GACTGGAAACTGCTGACGGAC 3'(position 536-557, relative to the A in the first codon of the human *SOX3* gene)

R26: 5' TGGCAGGTACATGCTGATCAGGTC 3'(1219-1243, relative to the A in the first codon of the human *SOX3* gene)

PCR reactions were performed in a 25-μl reaction volume containing 14 pM of forward and reverse primers, 7.5 mM dNTP, 1 x Thermo Pol buffer (New England Biolabs), 4 mM $MgSO_4$, 0.2 U Vent DNA polymerase (New England Biolabs), and 100 ng template genomic DNA. PCR amplification was performed with an initial denaturation of 30 s at 98°C; this was followed by 35 cycles of 98°C for 1 min, annealing/elongation at 73°C for 2 min, and a final extension at 72°C for 7 min.

The amplification product was electrophoresed in 1,5% agarose gel, purified using the QIAex II Gel Extraction Kit (Qiagen), and sequenced with the BigDye Terminator kit v 3.1 (Applied Biosystems) on the 3130 Genetic Analyzer (Applied Biosystems).

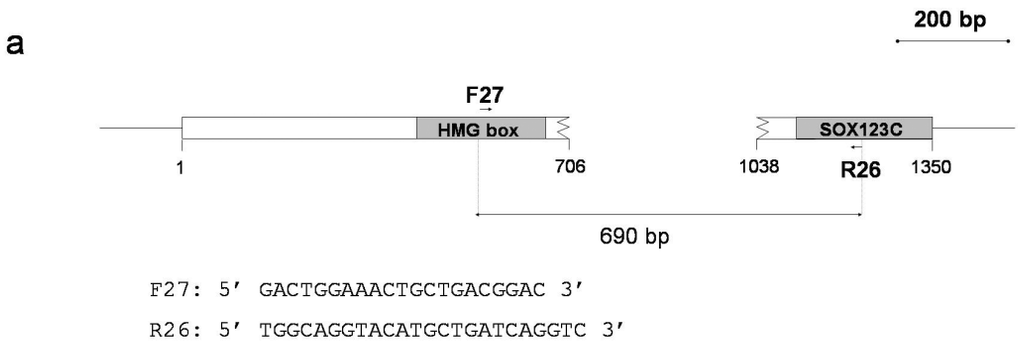
The nucleotide sequence data reported in this paper have been submitted to the GenBank database and assigned the accession number EU862290.

Analyses of deduced amino acid sequence of rat SOX3 protein

Prediction of the rat *Sox3* coding region, translation into an amino acid sequence, and amino acid composition analysis were performed with the The Sequence Manipulation Suite (http://www.bioinformatics.org/sms2/index.html). Domain search was performed using the ScanProsite tool (http://www.expasy.ch/prosite/). Database search and amino acid sequence alignment were performed using NCBI and ClustalW software.

RESULTS

The BLAST program using the human *SOX3* gene sequence (AL121875.10) as a query sequence revealed that the rat *Sox3* gene fragment was located within the rat chromosome X genomic contig based on Celera assembly NW_001091925.1. A partial coding sequence of rat *Sox3* was derived from nucleotide positions 3761772 – 3761469 and 3761153 – 3760451 (minus strand), with a sequence gap at position 3761469 – 3761153. In order to obtain the



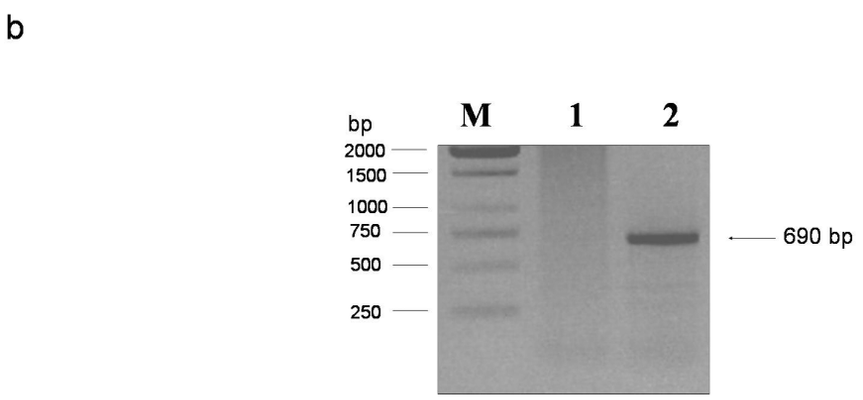


Fig. 1. a) Schematic presentation of the predicted rat Sox3 gene. The highly conserved HMG box and SOX123C domains are indicated as shaded boxes. Primer sequences are indicated, and their positions are presented by arrows. Predicted length of the PCR product is designated. b) PCR amplification on rat genomic DNA. M-DNA ladder, 1-negative control, 2-690 bp product generated by amplification of the rat *Sox3* gene.

complete rat *Sox3* coding sequence and to validate the assembly, we performed PCR amplification. PCR primers for rat *Sox3* gene amplification were designed to match regions that surround the gap and are conserved between the rat and human sequences (Fig. 1a). They were chosen to amplify across the missing sequence and to fill the existing gap in the rat *Sox3* sequence. The forward primer (F27) has a recognition sequence within the highly conserved HMG-box, while the reverse primer (R26) is encompassed within the sequence corresponding to the C-terminal region conserved among SOX1, SOX2, and SOX3 orthologs known as the SOX123C domain.

With rat genomic DNA as template, using primers F27 and R26 we obtained a fragment 690 bp in size (Fig. 1b). The amplified fragment was purified and sequenced using primers F27 and R26. Comparison of the obtained and rat contig sequences revealed identity over the regions between the primers and segments surrounding the gap (data not shown), indicating that a fragment corresponding to the rat *Sox3* gene was amplified. A composite sequence 1350 bp in size representing the complete rat *Sox3* coding sequence was assembled using the sequence obtained in this report (EU862290) and the rat genomic contig containing the partial sequence of this gene NW_001091925.1 (Fig. 2). Like other *Sox3* genes in vertebrates, the rat *Sox3* gene is encoded from a single exon, and no intron/exon junction sequence motifs are detected. Moreover, this analysis confirmed that, as in other mammals, *Sox3* is linked to the X-chromosome in rat.

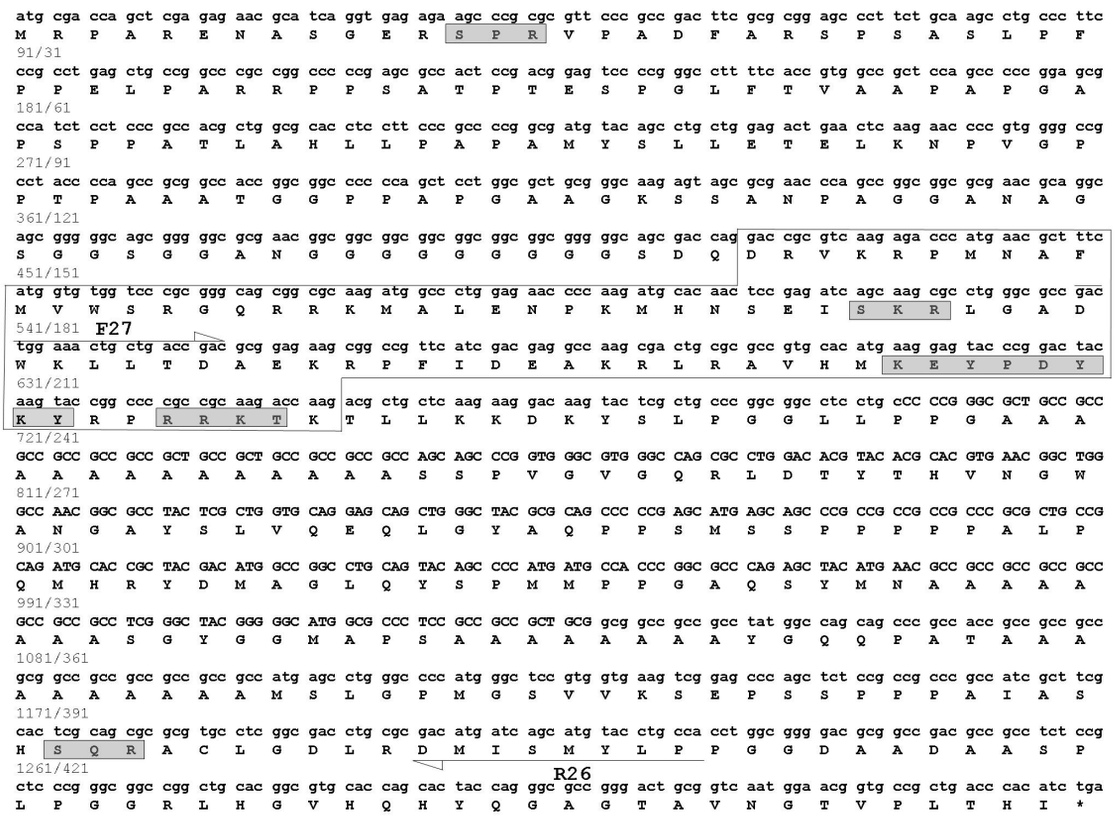


Fig. 2. Nucleotide and deduced amino acid sequence of the rat Sox3 gene. Incomplete genomic sequence from rat contig NW_001091925.1 is presented in lower case letters. Upper case letters present the sequence obtained in this study (EU862290). Arrows indicate primers used for PCR amplification. The HMG-box is boxed. Position of the stop codon is presented by an asterisk. Putative protein kinase C phosphorylation sites (SpR, SkR, and SqR, positions 13-15, 174-176, and 392-394, respectively), a tyrosine kinase phosphorylation site (KeypDyk.Y, position 205-212), and a cAMP- and cGMP-dependent protein kinase phosphorylation site (RRkT, position 215-218) are indicated as shaded boxes. Positions are given in relation to the first codon (methionine).

The deduced rat SOX3 amino acid sequence was obtained using the SMS Suite (Fig 2).

The 449 aa protein encoded by the rat *Sox3* gene is particularly rich in alanine (19.78%), glycine (11.33%), and proline (13.11%). Several putative sites for protein covalent modifications are revealed, viz., protein kinase C phosphorylation sites, a tyrosine kinase phosphorylation site, and a cAMP- and cGMP-dependent protein kinase phosphorylation site (Fig. 2).

The deduced amino acid sequence of rat Sox3 protein was compared with the GenBank sequence of human SOX3 protein (AL121875) (Fig. 3). Overall sequence identity was 93%. Moreover, domains characteristic for SOX B group proteins, including the HMG box (amino acid 141 to 219), group B homology (amino acid 220 to 227), and SOX123C domain (amino acid 371 to 449) are found to be identical in human and mouse SOX3 protein. This is in agreement with previous results obtained by comparative analyses of SOX3 orthologous proteins (K a t o h and K a t o h , 2005). Since humans and the common rat-mouse ancestor have been split for approximately 80 million years, the high conservation between human and rat SOX3 protein implies that this protein has been under strong evolution pressure, during which it has retained its functional properties.

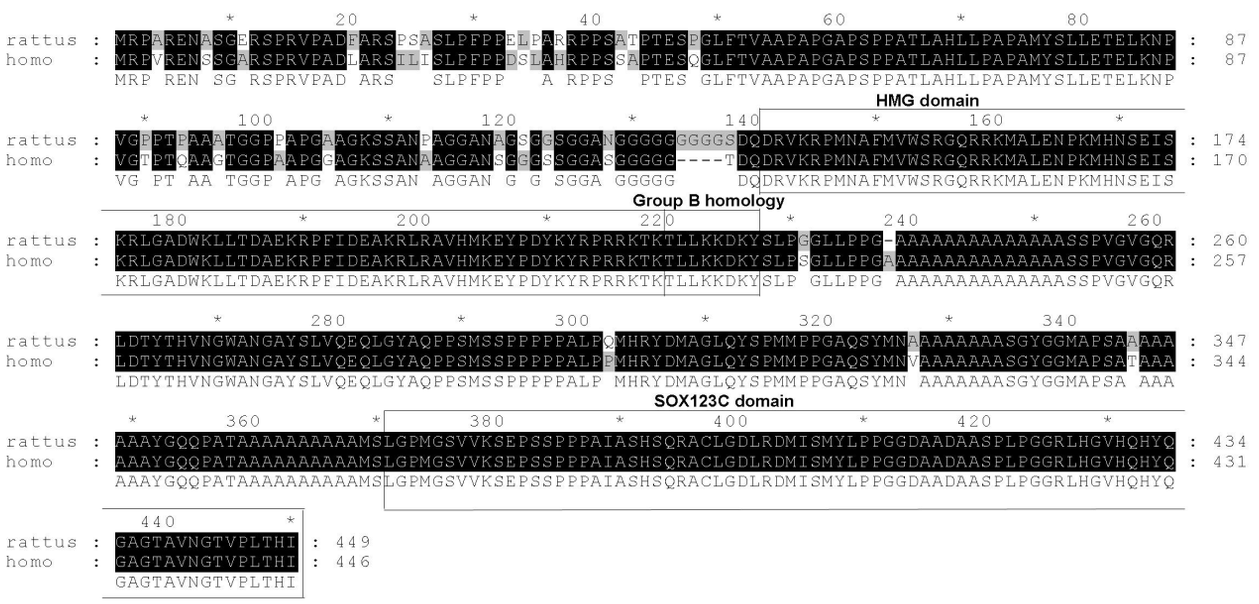


Fig. 3. Alignment of rat and human SOX3 proteins. Conserved amino acid residues are shown underneath the sequence. *Rattus: Rattus norvegicus; Homo: Homo sapiens.* Highly conserved regions, the HMG box, the Group B homology, and the SOX123C domain are boxed.

DISCUSSION

In this report, we have identified for the first time the rat *Sox3* gene using bioinformatic tools and sequencing. The complete coding sequence of the *Sox3* gene was determined by assembling the previously identified sequence within the genomic contig of rat chromosome Xq36 and sequencing data generated in this report. In addition, the deduced amino acid sequence of rat SOX3 protein was analyzed *in silico*. Rat SOX3 protein (449 aa) displays high conservation compared to its human orholog, indicating the importance of this protein in vertebrate development.

The laboratory rat (*Rattus norvegicus*) was the first mammalian species domesticated for scientific research, and it has become the most widely studied experimental animal model for biomedical research (Lindsey, 1979). Compared to the mouse, which is commonly considered as a genetic "model" organism, the rat has been better characterized physiologically. Accordingly, identification of novel rat genes provides the opportunity to elucidate their functions based on valuable phenotypic data already generated for this model organism (Jacob, 1999). The most convincing argument for comparative analysis of novel genes is that use of the three most studied mammalian species makes it possible to integrate genetics (mouse model), physiology (rat model), and clinical medicine.

The advance in knockout technology in rat has revealed striking differences in phenotypes compared to the corresponding mouse knockouts (Smits and Cuppen, 2006). More importantly, rat knockouts seem to reflect human biology better than mouse knockouts, indicating that rat models are more valuable than existing model organisms. Here reported for the first time, identification and characterization of the rat *Sox3* gene lays the foundation for *Sox3* knockout generation in rat, which would provide a valuable tool for better understanding the role of this gene during development and in adult life.

Acknowledgments — This work was supported by the Ministry of Science, Republic of Serbia (Grant No. 143028).

REFERENCES

Bowles, J., Schepers, G., and *P. Koopman* (2000). Phylogeny of the SOX family of developmental transcription factors

based on sequence and structural indicators. *Dev. Biol.* **227**, 239-255.

Brunelli, S., Silva Casey, E., Bell, D., Harland, R., and *R. Lovell-Badge* (2003). Expression of *Sox3* throughout the developing central nervous system is dependent on the combined action of discrete, evolutionarily conserved regulatory elements. *Genesis* **36**, 12-24.

Dong, C., Wilhelm, D., and *P. Koopman* (2004). Sox genes and cancer. *Cytogenet. Genome Res.* **105**, 442-447.

Gure, A. O., Stockert, E., Scanlan, M. J., Keresztes, R. S., Jager, D., and *N. K. Altorki* (2000). Serological identification of embryonic neural proteins as highly immunogenic tumor antigens in small cell lung cancer. *Proc. Natl. Acad. Sci. USA* **97**, 4198-4203.

Jacob, H. J. (1999). Functional genomics and rat models. *Genome Res.* **9**, 1013-1016.

Katoh, Y., and *M. Katoh* (2005). Comparative genomics on SOX2 orthologs. *Oncol. Rep.* **14**, 797-800.

Kim, R., Trubetskoy, A., Suzuki, T., Jenkins, N. A., Copeland, N. G., and *J. Lenz* (2003). Genome-based identification of cancer genes by proviral tagging in mouse retrovirus-induced T-cell lymphomas. *J. Virol.* **77**, 2056-2062.

Laumonnier, F., Ronce, N., Hamel, B. C., Thomas, P., Lespinasse, J., Raynaud, M., et al. (2002). Transcription factor SOX3 is involved in X-linked mental retardation with growth hormone deficiency. *Am. J. Hum. Genet.* **71**, 1450-1455.

Lindsey, J. (1979). Historical foundation in the laboratory rat, In: *The Laboratory Rat (*Eds. H. J. Baker, J. R. Lindsey, and S. H. Weisbroth), 1-36. Academic Press, New York.

Pevny, L. H., and *R. Lovell-Badge* (1997). Sox genes find their feet. *Curr. Opin. Genet. Dev.* **7**, 338-344.

Rizzoti, K., Brunelli, S., Carmignac, D., Thomas, P. Q., Robinson, I. C., and *R. Lovell-Badge* (2004). SOX3 is required during the formation of the hypothalamo-pituitary axis. *Nat. Genet.* **36**, 247-255.

Sasai, Y. (2001). Roles of Sox factors in neural determination: conserved signaling in evolution? *Int. J. Dev. Biol.* **45**, 321-326.

Smits, B. M., and *E. Cuppen* (2006). Rats go genomic. *Genome Biol.* **7**, 306.

Stankiewicz, P., Thiele, H., Schlicker, M., Cseke-Friedrich, A., Bartel-Friedrich, S., Yatsenko, S. A., et al. (2005). Duplication of Xq26.2-q27.1, including *SOX3*, in a mother and daughter with short stature and dyslalia. *Am. J. Med. Genet. A.* **138**, 11-17.

Stevanovic, M., Lovell-Badge, R., Collignon, J., and *P. N. Goodfellow* (1993). *SOX3* is an X-linked gene related to *SRY*. *Hum. Mol. Genet.* **2**, 2013-2018.

Uchikawa, M., Kamachi, Y., and H. *Kondoh* (1999). Two distinct subgroups of Group B *Sox* genes for transcriptional activators and repressors: their expression during embryonic organogenesis of the chicken. *Mech. Dev.* **84**, 103-120.

Wegner, M. (1999). From head to toes: the multiple facets of *Sox* proteins. *Nucleic Acids Res.* **27**, 1409-1420.

Weiss, J., Meeks, J. J., Hurley, L., Raverot, G., Frassetto, A., and *J. L. Jameson* (2003). *Sox3* is required for gonadal function, but not sex determination, in males and females. *Mol. Cell. Biol.* **23**, 8084-8091.

Woods, K. S., Cundall, M., Turton, J., Rizotti, K., Mehta, A., and *R. Palmer* (2005). Over- and underdosage of SOX3 is associated with infundibular hypoplasia and hypopituitarism. *Am. J. Hum. Genet.* **76**, 833-849.

Xia, Y., Papalopulu, N., Vogt, P. K., and *J. Li* (2000). The oncogenic potential of the high mobility group box protein Sox3. *Cancer Res.* **60**, 6303-6306.

АМПЛИФИКАЦИЈА И АНАЛИЗА СЕКВЕНЦЕ *SOX3* ГЕНА ПАЦОВА

А. КРСТИЋ, МАРИЈА МОЈСИН, НАТАША КОВАЧЕВИЋ-ГРУЈИЧИЋ и МИЛЕНА СТЕВАНОВИЋ

Институт за молекуларну генетику и генетичко инжењерство, 11010 Београд, Србија

Sox3 ген је један од маркера најранијих фаза развића нервног система кичмењака који је укључен у контролу диференцијације нервних прекурсора. Упркос чињеници да је геном пацова секвенциран и јавно доступан, само парцијална секвенца *Sox3* гена ове врсте је била депонована у бази података. У овом раду смо применом PCR-а, секвенцирања и биоинформатичке анализе генерисали комплетну кодирајућу секвенцу *Sox3* гена пацова. Анализа добијене секвенце је показала да *Sox3* ген кодира протеин од 449 амино киселина. Упоредна анализа ортологих SOX3 протеина пацова и човека показала је висок степен еволутивне очуваности. Идентификација и карактеризација *Sox3* гена пацова допринеће бољем разумевању његове улоге током развића нервног система и омогућиће бољи увид у еволуцију овог гена код вертебрата.