

GENETIC ANALYSIS OF SUPEROXIDE DISMUTASE 1 GENE IN MURRAH RIVER BUFFALO

N. B. Stafuzza, M. M. Borges and M. E. J. Amaral-Trusty*

Department of Biology, Instituto de Biociências, Letras e Ciências Exatas (IBILCE)/São Paulo State University (UNESP), São José do Rio Preto-SP, 15054-000, Brazil

*Corresponding author e-mail: eamaral@ibilce.unesp.br

ABSTRACT

In this study we characterized the complete sequence of the SOD1 gene from an animal representing the Murrah breed of the river buffalo (*Bubalus bubalis*) and we compared its coding sequence and the amino acid sequence with its homologous from others mammals. The buffalo SOD1 gene contains 8,720 bp in length organized into five exons (72 bp, 91 bp, 70 bp, 118 bp and 108 bp, respectively), four introns (3,630 bp, 1,768 bp, 827 bp and 1,614 bp, respectively), 5'UTR (103 bp) and 3'UTR (319 bp), with the exon/intron size ratio of 1:17.07 and the CG level of 46.42 %. A total of 12 repetitive elements were identified at intronic level. Comparative analysis between SOD1 gene coding sequence and the amino acid sequence with its homologous from other mammalian species showed a percentage identity varying from 82% to 98% at DNA coding level and 81% to 97% at amino acids level. In addition, the alignment of the complete SOD1 gene sequence between the Murrah and the Mediterranean breeds revealed nine potential SNPs which could be candidates for validation in commercial buffalo populations.

Key words: BAC library, *Bubalus bubalis*, Murrah breed, Pyrosequencing, SOD1.

INTRODUCTION

In the last decade, the research community has generated buffalo (*Bubalus bubalis*) genome resources through whole genome mapping (Amaral *et al.*, 2007; 2008), whole genome sequencing (Murrah breed) (Tantia *et al.*, 2011) and the construction of a buffalo BAC library (Stafuzza *et al.*, 2012) especially useful for region-specific resequencing. Earlier this year, a buffalo whole genome sequence (Mediterranean breed) generated by next-generation sequencing platforms was publicly released at the National Center for Biotechnology Information database (<http://www.ncbi.nlm.nih.gov/assembly/67671>).

Although the next-generation sequencing platforms has been used to produce genome sequences for two buffalo breeds, recent studies of de novo assemblies produced by these approaches have indicated a loss of approximately 16% of the genome (Alkan *et al.*, 2011).

Large-insert clone libraries, such as the buffalo BAC library (Stafuzza *et al.*, 2012) are still required for validation and accurate assembly of the genome sequence, allowing for instance isolation of genes and gene clusters, contributing to the elucidation of gene organization and identification of regulatory and repetitive elements.

Superoxide dismutases (SOD) are the major antioxidant enzymes responsible for protecting and maintaining the antioxidative/oxidative balance in different cell compartments and in the extracellular space

(McCord and Fridovich, 1988; Wallace and Melov, 1998). Three superoxide dismutases enzymes have been identified and characterized in mammals: cooper/zinc superoxide dismutase (Cu/ZnSOD, encoded by SOD1 gene), manganese superoxide dismutase (MnSOD, encoded by SOD2 gene) and the extracellular superoxide dismutase (ECSOD, encoded by SOD3 gene) (Fridovich, 1995; Miao and Clair, 2009).

Studies in cattle (Larska *et al.*, 2010) and sheep (Garcia-Crespo *et al.*, 2006) with transmissible spongiform encephalopathy (TSE) have identified SOD1 gene as a candidate for prion diseases. This gene also plays an important role in mammalian reproduction maintaining the fertility and promoting the development of embryo in cattle (Nakamura *et al.*, 2001; Valdez *et al.*, 2005; Lee *et al.*, 2010; Vu *et al.*, 2012; Kankofer *et al.*, 2013), sheep (Al-Gubory *et al.*, 2003, 2004, 2005) and mice (Foyouzi *et al.*, 2005; Noda *et al.*, 2012).

SOD1 gene has been previously mapped in the river buffalo chromosome 1 by two independent methods, cytogenetic (Iannuzzi *et al.*, 2003) and radiation hybrid mapping (Miziara *et al.*, 2007; Amaral *et al.*, 2008).

In this study, we characterized a genomic region of 44.79 kb from buffalo chromosome 1 containing the SOD1 gene, by sequencing a clone from the buffalo BAC library. The SOD1 molecular structure was determined and a comparative analysis was performed using Murrah and Mediterranean buffalo breeds to search for potential nucleotide variations.

MATERIALS AND METHODS

Bacterial artificial chromosome library screening: The buffalo BAC library (Stafuzza *et al.*, 2012) was 3D screened by PCR with the same cattle-derived primers used on RH mapping of SOD1 gene in the buffalo chromosome 1 (forward 5'-GTTTGGCCTGTGGTGAATTGGAA-3' and reverse 5'-GGCCAAAATACAGAGATGAATGAA-3') (Barendse *et al.*, 1994).

The 3D screening of the library was carried out in three steps. The first step was performed with 17 superpools where each superpool contained BAC clones from eight 384-well plates. In the second step, eight single pools, representing BAC clones from each of the eight plates in the positive superpool, were screened. Finally, the pooled row and column BAC clones from the positive plate were screened. The intersection of the row and column identified the location of the positive BAC clone for the SOD1 gene.

Polymerase chain reactions included: 10 mM Tris-HCl, 1.5 mM MgCl₂, 50 mM KCl, 10 mM dNTPs, 0.2 mM of each primer, 0.5 unit of AmpliTaq Gold polymerase (Life Technologies™, USA) and 2 µL of BAC DNA in a 25 µL reactions volume. The amplification conditions were as follows: initial denaturation at 94°C for 10 min, followed by 35 cycles at 94°C for 30 s (denaturation), 50°C for 30 s (annealing), 72°C for 30 s (extension) and a final extension at 72°C for 7 min. The PCR products were electrophoresed through 2% agarose gels in 1X TBE buffer containing 0.25 µg/ml ethidium bromide and photographed under UV light.

Isolation of Murrah BAC DNA: The positive clone for SOD1 gene was grown in 50 mL of Luria-Bertani medium (Sigma-Aldrich, USA) containing 12.5 µg/mL chloramphenicol. The BAC DNA was purified using the PhasePrep™BAC DNA Kit (Sigma-Aldrich, USA) as described by the manufacturer. Briefly, clones were harvested by centrifugation at 5,000 × *g* for 10 min. Supernatant was removed and the pellet was resuspended in 250 µL of a solution containing RNase and 250 µL of lysis components followed by the addition of 250 µL of neutralization solution followed by incubation on ice for 5 min and centrifugation at 16,000 × *g* for 5 min at 4°C. The supernatant was transferred to a 2 mL microcentrifuge tubes and 450 µL of isopropanol was added. The nucleic acids were collected by centrifugation at 16,000 × *g* for 20 min at 4°C. The pellet was washed with 100 µL of 70% ethanol and followed by addition of 500 µL of elution solution.

The endotoxins and other impurities were removed by addition of 100 µL of endotoxin removal solution followed by incubation on ice for 5 min and 37°C for 5 min. The clear upper phase containing the

BAC DNA was transferred to another 2 mL microcentrifuge tube follow by adding 540 µL of DNA precipitation solution and centrifugation at 16,000 × *g* for 20 min at 4°C. The supernatant was removed and the pellet was washed with 150 µL of 70% ethanol followed by centrifugation at 16,000 × *g* for 20 min at 4°C. The supernatant was discarded and the BAC DNA was eluted in 100 µL of TE buffer.

The DNA concentration was determined by NanoDrop spectrophotometer (Thermo Scientific, USA) and adjusted to 100 ng/µL. An additional PCR reaction with the purified DNA was performed as described earlier to confirm the amplification of the SOD1 gene.

Murrah BAC DNA sequencing and bioinformatics analysis: The BAC DNA sequencing was obtained by Next Generation Sequencing (NGS) using 454-pyrosequencing - GS FLX Titanium chemistry (Roche) performed at 454 Sequencing Center, Branford, CT, USA. The sequence data was assembled using the GS De Novo Assembler version 2.6 (Roche).

Repetitive DNA elements in the Murrah BAC clone sequence were identified by RepeatMasker (<http://www.repeatmasker.org/>). The software Augustus was used to predict genes from the masked nucleotide sequence (<http://bioinf.uni-greifswald.de/augustus/>). The predicted gene sequence was aligned against NCBI reference mRNA sequences from *Bos taurus* using BLAST tools (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) and the predicted peptide sequence was aligned against non-redundant protein sequences by BLASTp (protein-protein BLAST) to verify homology among other species and identify the gene predicted.

The coordinates of exons, introns, 5'UTR and 3'UTR were obtained performing an alignment between the BAC clone sequence from the Murrah breed against the predicted gene sequence available for the Mediterranean breed on NCBI (<http://www.ncbi.nlm.nih.gov/assembly/67671>). In addition, the Splign program (<http://www.ncbi.nlm.nih.gov/sutils/splign/splign.cgi>) was used to recognizing introns and splice signals.

VISTA plots were generated using wgVISTA alignment (<http://genome.lbl.gov/cgi-bin/WGVistaInput>) with the bovine chromosome 1 genomic scaffold (GenBank accession no NW_003103793.1) to show the relative nucleotide conservation between buffalo and bovine.

The Clustal Omega software (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) was used to align coding gene sequences and amino acids sequences from buffalo (*Bubalus bubalis*), cattle (*Bos taurus*), goat (*Capra hircus*), sheep (*Ovis aries*), horse (*Equus caballus*), pig (*Sus scrofa*), human (*Homo sapiens*), mouse (*Mus musculus*) and rat (*Rattus norvegicus*).

RESULTS AND DISCUSSION

The screening of the buffalo BAC library resulted in the identification of one positive clone for the SOD1 gene. The 454/Roche GS FLX sequencer generated a total of 3,991 reads with an average read length of 697 nucleotides arranged in one contig of 44,794 nucleotides providing a 60.38X coverage, N50 value of 44.79 Kb and 99.91% of Q40+. The Murrah BAC DNA sequence was deposited on GenBank under accession number KJ635889.

The gene prediction by Augustus software detected one gene in the Murrah BAC sequence. The alignment of the predicted gene against the NCBI reference mRNA sequences from *Bos taurus* revealed 98% of identity and 62% of coverage with bovine superoxide dismutase 1, soluble (SOD1) mRNA (GenBank accession no XM_005201085.1). The predicted peptide alignment against non-redundant protein sequences from *Bos taurus* database revealed 94% of identity and 65% of coverage with bovine cooper/zinc superoxide dismutase protein (GenBank accession no NP_777040.1).

The alignment between the Murrah BAC sequence against the predicted gene for the Mediterranean breed (GenBank accession no NW_005784534.1, unplaced genomic scaffold scf7180021616654, UMD_CASPUR_WB_2.0), established the coordinates of the SOD1 gene from nucleotide 6,530 to nucleotide 15,249, with a total length of 8,720 bp. The organization of the gene was as following: 5'UTR (103 bp), exon 1 (72 bp), intron 1 (3,630 bp), exon 2 (91 bp), intron 2 (1,768 bp), exon 3 (70 bp), intron 3 (827 bp), exon 4 (118 bp), intron 4 (1,614 bp), exon 5 (108 bp) and 3'UTR (319 bp), with the exon/intron size ratio of 1:17.07 and the CG level of 46.42 %. The SOD1 coding sequence showed a total of 459 nucleotides which encodes the cooper/zinc superoxide dismutase protein with 152 amino acids, the same sizes observed in others bovids (cattle, sheep and goat), but smaller than sizes observed to pig (462 nucleotides and 153 amino acids), human, horse and rodents (465 nucleotides and 154 amino acids).

The alignment of the SOD1 gene sequence from Murrah breed against the predicted gene sequence from the Mediterranean breed revealed 99% of identity. A total of nine nucleotide substitutions, six insertions and two deletions were observed in the Murrah breed sequence with the positions as following: intron 1 (g.419_420insG, g.432G>A, g.433T>C, g.467_468insG, g.477_478insC, g.584A>G, g.1,587A>G, g.2,090C>G, g.3,031_3,032insGTT and g.3,411delT), intron 2 (g.5,650delC), intron 3 (g.5,822G>A) and intron 4 (g.6,748A>G, g.7,539T>C and g.7,902A>G). These substitutions are potential SNPs (Single Nucleotide Polymorphisms) that could be used to

uncover the genetic structure of these two buffalo breeds in future population studies.

The comparative analysis using the wgVISTA alignment between the SOD1 gene from buffalo and the bovine (chromosome 1 genomic scaffold - GenBank accession no NW_003103793.1) showed a percentage of genomic conservation of 91.7% on the 5'UTR region, 98.7% on exon 1, 94.9% on intron 1, 97.8% on exon 2, 94.1% on intron 2, 100% on exon 3, 97.1% on intron 3, 96.6% on exon 4, 95.1% on intron 4, 99.1% on exon 5 and 96.9% on 3'UTR (Fig. 1).

The Repeat Masker tool identified a CG content of 45% and a total of 113 repetitive elements in the Murrah BAC DNA sequence. Statistical analysis revealed that 51.53% of the buffalo DNA (23,083 bp) comprises interspersed repeats, represented by 46 small interspersed nucleotide elements (SINEs), 18 long interspersed nucleotide elements (LINEs), 17 long terminal repeat elements (LTR) and 9 DNA elements. Others repetitive elements include 8 small RNAs, 10 simple tandem repeats, 3 satellites and 2 low-complexity DNA. The repetitive elements are summarized on Table 1.

Regarding the presence of repetitive elements in the buffalo SOD1 gene, the RepeatMasker tool identified a total of 12 repetitive elements at intronic level including seven SINEs, two DNA elements, two simple tandem repeats and one low-complexity DNA element, corresponding to 18.27% of the total length of gene (Fig. 2). Comparative analysis including other bovids (cattle, sheep and goat) revealed the G-rich low-complexity DNA element specific in the buffalo intron 1, while the others elements are common with other bovids.

The multiple alignment of the buffalo SOD1 coding sequence showed 98% of identity with cattle (GenBank accession no NM_174615.2), 95% with goat (GenBank accession no NM_001285550.1), 94% with sheep (GenBank accession no XM_004012626.1), 88% with pig (GenBank accession no NM_001190422.1), 85% with human (GenBank accession no NM_000454.4), 83% with horse and mouse (GenBank accession no NM_001081826.2 and NM_011434.1, respectively) and 82% with rat (GenBank accession no NM_017050.1). All these alignments showed 100% of coverage, except with horse that showed 98% of coverage (Fig. 3). Miao and Clair (2009) already had observed the significant similarity existent among rat, mouse, human and bovine SOD 1 gene in terms of size of exons, particularly the coding regions.

On the protein level, the predicted amino acid sequence of the buffalo cooper/zinc superoxide dismutase protein revealed 97% of identity with cattle (GenBank accession no NP_777040.1), 95% with goat (GenBank accession no NP_001272479.1), 94% with sheep (GenBank accession no XP_004012675.1), 85% with pig (GenBank accession no NP_001177351.1), 83% with mouse (GenBank accession no NP_035564.1), 82% with

rat (GenBank accession no NP_058746.1) and 81% with horse and human (GenBank accession no NP_001075295.1 and NP_000445.1, respectively). The alignments with goat, sheep and horse showed 99% of coverage, while the alignments with cattle, pig, human e rodents showed 100% of coverage (Fig. 4).

The differences observed between coding sequences of buffalo and cattle revealed a total of ten nucleotide substitutions being six synonymous and four nonsynonymous resulting in the following amino acids substitutions: Val6Ile, Asp26Asn, Pro91Ser and Glu121Asp. These four amino acids substitutions are responsible for the identity of 97% observed between the

buffalo and cattle cooper/zinc superoxide dismutase protein. Due the buffalo SOD1 coding sequence and amino acid sequence showed the highest identity with cattle, we suggest that this protein in buffalo is more similar functionally to cattle.

The comparative analysis of the buffalo SOD1 gene confirmed the highly conservation of its structure when compared with several mammals, providing basic information for future studies of this gene in different buffalo herds involving polymorphisms association with economically important traits, especially those related with fertility and development of neurodegenerative diseases.

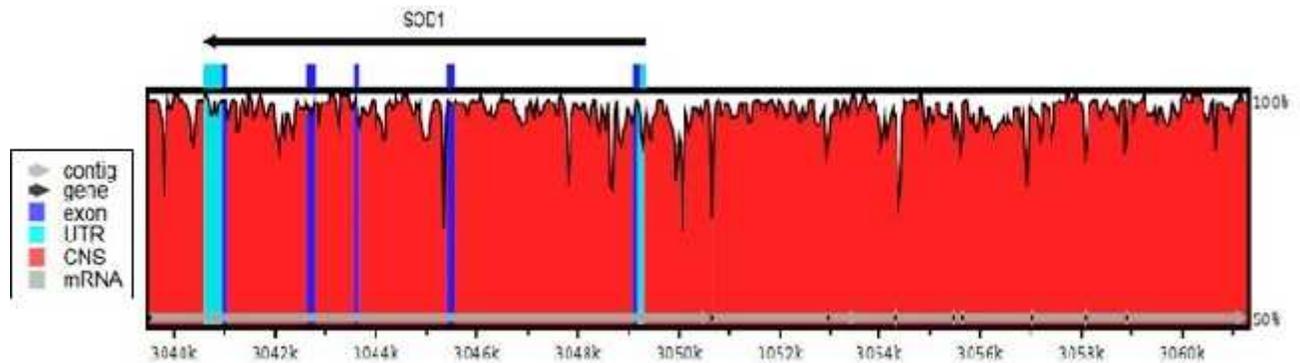


Fig. 1. wgVISTA plot obtained from the alignment between the SOD1 gene from buffalo and bovine (chromosome 1 genomic scaffold - GenBank accession No NW_003103793.1) showing a genomic conservation (CNS) higher than 91% between both sequences, indicated by red color. The black arrow represents the SOD1 gene. Exons are indicate by dark blue blocks while the 5'UTR and 3'UTR are indicated by the light blue blocks.

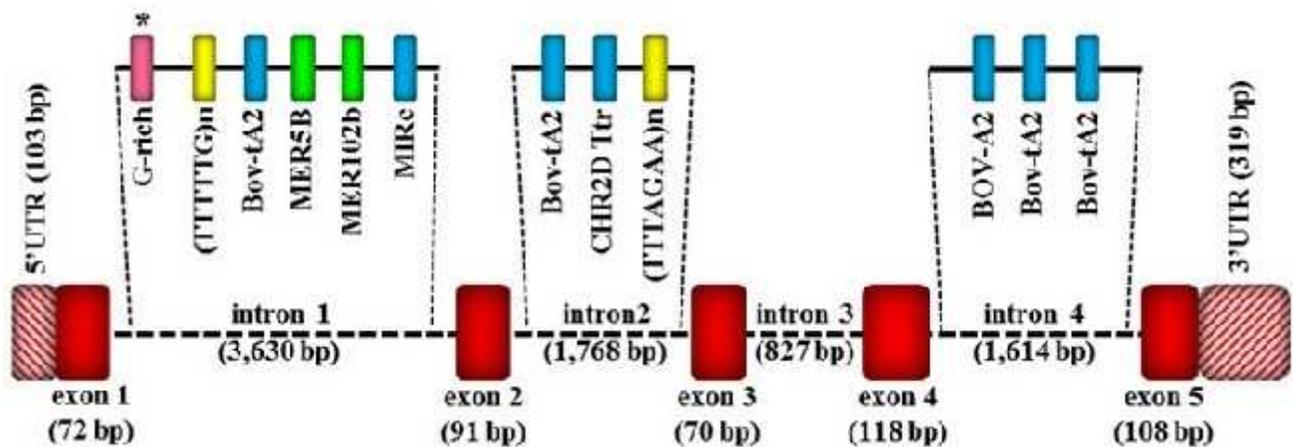


Fig. 2. Schematic representation of the buffalo SOD1 gene showing a total of 12 repetitive elements found in the introns. The repetitive elements are indicated by the color boxes delimited by the size of each intron. SINEs are represented by blue boxes, DNA elements (green), simple tandem repeats (yellow) and low-complexity DNA (pink). The asterisk highlight the buffalo specific G-rich repetitive element found on intron 1. The red boxes represent exons 1, 2, 3, 4 and 5 and the hachured boxes represent the 5'UTR and 3'UTR, with their respective sizes in bp (base pairs).

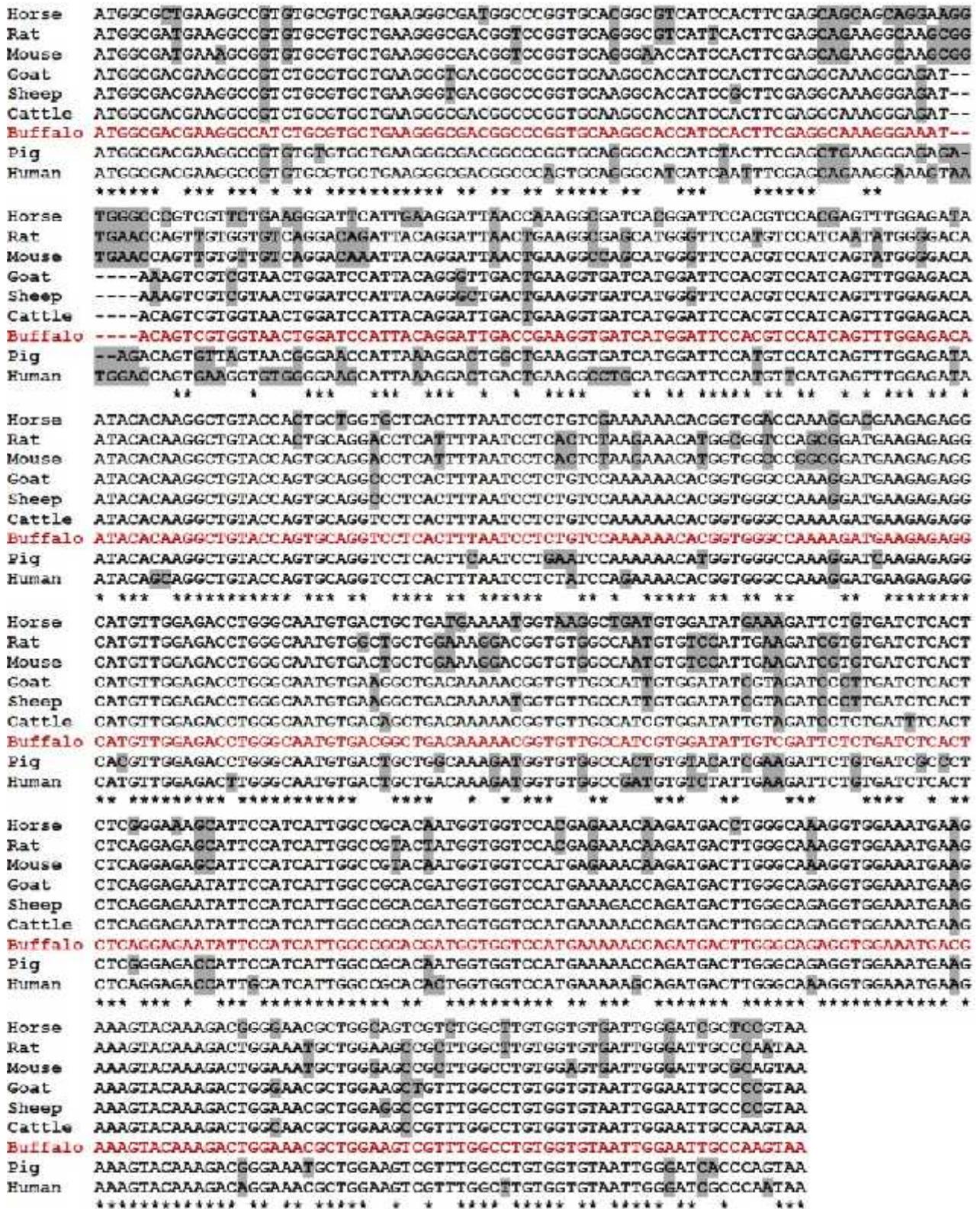


Fig. 3. Multispecies alignment with Clustal Omega software showing the percentage of the identity among the buffalo SOD1 gene coding sequence and horse, rat, mouse, goat, sheep, cattle, pig and human. Nucleotides substitutions in relation to buffalo gene coding sequence are highlighted in gray.

- Greco, J. Jeong, Z. Cai, G. Lin, A. Prasad, S. Kumar, G.P. Saradhi, B. Mathew, M.A. Kumar, M.N. Mizziara, P. Mariani, A.R. Caetano, S.R. Galvão, M.S. Tantia, R.K. Vijh, B. Mishra, S.T. Kumar, V.A. Pelai, A.M. Santana, L.C. Fornitano, B.C. Jones, H. Tonhati, S. Moore, P. Stothard and J.E. Womack (2008). A first generation whole genome RH map of the river buffalo with comparison to domestic cattle. *BMC Genomics* 9: 631.
- Barendse, W., S.M. Armitage, L.M. Kossarek, A. Shalom, B.W. Kirkpatrick, A.M. Ryan, D. Clayton, L. Li, H.L. Neibergs, N. Zhang, W.M. Grosse, J. Weiss, P. Creighton, F. McCarthy, M. Ron, A.J. Teale, R. Fries, R.A. McGraw, S.S. Moore, M. Georges, M. Soller, J.E. Womack and D.J.S. Hetzel (1994). A genetic linkage map of the bovine genome. *Nat. Genet.* 6(3): 227-235.
- Foyouzi, N., Z. Cai, Y. Sugimoto and C. Stocco (2005). Changes in the expression of steroidogenic and antioxidant genes in the mouse corpus luteum during luteolysis. *Biol. Reprod.* 72: 1134-1141.
- Fridovich, I. (1995). Superoxide radical and superoxide dismutases. *Annu. Rev. Biochem.* 64: 97-112.
- Garcia-Crespo, D., R.A. Juste and A. Hurtado (2006). Differential gene expression in central nervous system tissues of sheep with natural scrapie. *Brain Res.* 1073-1074: 88-92.
- Iannuzzi, L., G.P. Di Meo, A. Perucatti, L. Schibler, D. Incarnato, D. Gallagher, A. Eggen, L. Ferretti, E.P. Cribiu and J. Womack (2003). The river buffalo (*Bubalus bubalis*, 2n=50) cytogenetic map: assignment of 64 loci by fluorescence in situ hybridization and R-banding. *Cytogenet. Genome Res.* 102(1-4): 65-75.
- Kankofer, M., J. Wawrzykowski and M. Hoedemaker (2013). The presence of SOD1 and GSH-Px in bovine retained and properly released foetal membranes. *Reprod. Domest. Anim.* 48(4): 699-704.
- Larska, M., M.P. Polak, J.F. Zmudzinski and J.M. Torres (2010). Comparison of mRNA expression levels of selected genes in the brain stem of cattle naturally infected with classical and atypical BSE. *Brain Res.* 1351: 13-22.
- Lee, S., T.J. Acosta, Y. Nakagawa and K. Okuda (2010). Role of nitric oxide in the regulation of superoxide dismutase and prostaglandin F₂(alpha) production in bovine luteal endothelial cells. *J. Reprod. Dev.* 56(4): 454-459.
- McCord, J. M. and I. Fridovich (1988). Superoxide dismutase: the first twenty years (1968-1988). *Free Radic. Biol. Med.* 5(5-6): 363-369.
- Miao, L. and D.K.St. Clair (2009). Regulation of superoxide dismutase genes: implication in disease. *Free Radic. Biol. Med.* 47(4): 344-356.
- Mizziara, M.N., T. Goldammer, N.B. Stafuzza, P. Ianella, R. Agarwala, A.A. Schaffer, J.S. Elliott, P.K. Riggs, J.E. Womack and M.E. Amaral (2007). A radiation hybrid map of river buffalo (*Bubalus bubalis*) chromosome 1 (BBU1). *Cytogenet. Genome Res.* 119(1-2): 100-104.
- Nakamura, T. and K. Sakamoto (2001). Reactive oxygen species up-regulates cyclooxygenase-2, p53, and Bax mRNA expression in bovine luteal cells. *Biochem. Biophys. Res. Commun.* 284(1): 203-210.
- Noda, Y., K. Ota, T. Shirasawa and T. Shimizu (2012). Copper/zinc superoxide dismutase insufficiency impairs progesterone secretion and fertility in female mice. *Biol. Reprod.* 86: 1-8.
- Stafuzza, N.B., C.A. Abbey, C.A. Gill, J.E. Womack and M.E. Amaral (2012). Construction and preliminary characterization of a river buffalo bacterial artificial chromosome library. *Genet. Mol. Res.* 11(3): 3013-3019.
- Tantia, M.S., R.K. Vijh, V. Bhasin, P. Sikka, P.K. Vij, R.S. Kataria, B.P. Mishra, S.P. Yadav, A.K. Pandey, R.K. Sethi, B.K. Joshi, S.C. Gupta and K.M.L. Pathak (2011). Whole-genome sequence assembly of the water buffalo (*Bubalus bubalis*). *Indian J. Anim. Sci.* 81(5): 38-46.
- Valdez, K.E., S.P. Cuneo and A.M. Turzillo (2005). Regulation of apoptosis in the atresia of dominant bovine follicles of the first follicular wave following ovulation. *Reproduction.* 130(1): 71-81.
- Vu, H.V., S. Lee, T.J. Acosta, S. Yoshioka, H. Abe and K. Okuda (2012). Roles of prostaglandin F₂alpha and hydrogen peroxide in the regulation of Copper/Zinc superoxide dismutase in bovine corpus luteum and luteal endothelial cells. *Reprod. Biol. Endocrinol.* 10:87.
- Wallace, D.C. and S. Melov (1998). Radicals r'aging. *Nat. Genet.* 19(2): 105-106.