ISSN 1682-296X (Print)
ISSN 1682-2978 (Online)

Bio Technology

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Biotechnology 6 (3): 393-407, 2007
ISSN 1682-296X

Identification and Assignment of Human Superoxide Dismutase-1 Retropseudogenes Using Comparative Mapping and Bioinformatic Analysis

Ahlam A. Abou Mossallam
Department of Cell Biology, Genetic Engineering and Biotechnology Division,
National Research Center, Dokki, Giza 12622, Egypt

Abstract: In the course of studying the natural immunity of cattle and other rural ruminants in Egypt, SOD1 was one of the genes under investigation. The nucleotide sequence analysis of the obtained SOD1 amplified segments of cattle lead to the identification and assignment of SOD1 retropseudogenes Ψ 71.4 and 69.1 on human chromosome 8 and 16, respectively. Sequencing and Blast analysis for SOD1 amplified segments (370 nt) of Egyptian native cattle cDNA and genomic DNA revealed that these segments have 97% similarity. Blast analysis showed that the amplified sequence of cattle genomic DNA has great similarity with Bos taurus SOD-1 mRNA; with four separate segments of Bos taurus chromosome 1 genomic superoxide dismutase 1 and showed high identities with DNA of Bos taurus chromosome 13 similar to SOD1. These results indicate that cattle genomic DNA has intronless pseudogene on Bos taurus chromosome 13. The amplified sequence of cattle genomic DNA also showed a high similarity with three separate segments of human chromosome 21 on which the SOD1 gene is known to be located. In addition, four interesting alignments were also observed; it showed high alignment with a segment of human chromosome 16; a segment of human chromosome 8 and with human SOD1 processed pseudogenes Ψ 69.1 and 71.4. The SOD1 processed pseudogene Ψ 69.1 can be assigned to human chromosome 16 and the SOD1 processed pseudogene Ψ 71.4 can be assigned to chromosome 8. Two segments of 2000 nt from both human chromosome 16 and 8 were used for homology search. The maches obtained have high identities with the whole length of human SOD1 processed pseudogenes Ψ 69.1 and 71.4, respectively. These results identify the complete SOD-1 retropseudogenes sequence and their location on human chromosome 16 and 8, respectively. They bear the structural hallmarks of processed pseudogene; they lack introns; the 5' and 3' direct repeats are presented. They also contain TATA box upstream from the 5' direct repeat in a position almost identical to the functional TATA box. They also contain a 3' poly (A or T)-rich stretch and are flanked by short direct repeat sequences. The presence of SOD1 retropseudogenes on human chromosome 16 and 8 reported in this study explain the reported FALS and other diseases in individuals with a normal chromosome 21.

Key words: Human chromosome 16, human chromosome 8, FALS, SOD1, retropseudogene

INTRODUCTION

There are several isoforms of Superoxide Dismutase (SOD) that differ in metal binding ability, distribution in different cell compartments and in sensitivity to various reagents. Among these, Cu/Zn Superoxide Dismutase (SOD1) is widely distributed and comprises 90% of the total SOD (Noor *et al*., 2002). Bovine Cu/Zn superoxide dismutase gene was assigned to bovine chromosome 1 (Schmutz *et al*., 1996), whereas, bovine Cu/Zn SOD1 pseudogene was assigned to bovine chromosome 13 (Gallagher *et al*., 1999). In human, Cu/Zn superoxide dismutase gene was assigned to chromosome 21q in the segment enclosing the distal part 21q21-21q22.1 (Huret *et al*., 1987).

Noor *et al*. (2002) described the role of SODs, especially SOD1, in several diseases such as Familial Amyotrophic Lateral Sclerosis (FALS); Parkinson's disease; Alzheimer's disease; dengue fever; cancer; Down's syndrome; cataract and several neurological disorders. Mutations in the SOD1 gene have been found to lead to the development of FALS (Andersen *et al*., 2003; Hough *et al*., 2004).

However approximately 80% of FALS cases could not be related to SOD1 mutations; thus, other candidate genes have been tested for their possible role in disease pathogenesis. The other isoforms of SOD-that is, MnSOD (SOD2) and extracellular SOD (SOD3)-have been examined, but none could be linked to FALS (de Belleroche *et al*., 1995; Siddique and Deng, 1996).

The abundance of pseudogenes is a remarkable feature of mammalian genomes. Pseudogenes are copies of specific genes and are present in every mammalian chromosome (Gonçalves *et al*., 2000; Chen *et al*., 2002). In general, pseudogenes are thought to be nonfunctional (Mighell *et al*., 2000) as they have accumulated vast numbers of mutations during evolution and have lost the ability to be transcribed. Pseudogenes fall into two distinct categories depending on the mechanism by which they are generated: Processed pseudogenes are reverse transcribed from mRNAs (and thus do not contain introns) whereas nonprocessed pseudogenes arise from duplications of genomic DNA (Mighell *et al*., 2000; Harrison *et al*., 2002a). Among the abundant processed pseudogenes, there are a substantial number of processed gens or retrogenes of novel function that also derive from mRNAs of various intron-containing genes (Brosius, 1999; Lahn and Page, 1999; Betra`n *et al*., 2002). A number of retropseudogenes have also been implicated in various human diseases (Zhang *et al*., 2003). Parental genes of human processed pseudogenes are of various types, including those for enzymes, structural proteins and regulatory proteins such as ligand-binding proteins and transcription factors (Ohshima *et al*., 2003). It is unclear how many pseudogenes exist in the human genome. Estimates for the number of human genes range from ~22,000 to ~75,000 (Harrison *et al*., 2002b). It is thought that up to 22% of these gene predictions may be pseudogenic (Lander *et al*., 2001; Yeh *et al*., 2001).

It is important to characterize the human pseudogene population, as their existence interferes with gene identification and annotation. They are also an important resource for the study of the evolution of protein families. In the course of studying the natural immunity of cattle and other rural ruminants in Egypt, SOD1 was one of the genes under investigation. The nucleotide sequence analysis of the obtained SOD1 amplified segments of cattle lead to the identification and assignment of SOD1 retropseudogenes 71.4 and Ψ 69.1 on human chromosome 8 and 16.

MATERIALS AND METHODS

Collection of samples and genomic DNA extraction: Blood samples of Egyptian native cattle were collected in Ethylene Diamine Tetra Acetic Acid (EDTA) and were processed as soon as possible. Genomic DNA was extracted from leucocytes following the established protocols (Blin and Stafford, 1976) modified by Shih and Weinberg (1982). DNA was finally dissolved in an appropriate volume (200 μL) of 1x Tris-EDTA (TE). Using a spectrophotometer, the DNA concentration was determined and diluted in sterile water to a concentration of 50 ng μL^{-1}.

Total RNA extraction and reverse transcription: Total RNA from blood was extracted using Trifast reagent after RBCs hydrolysis. To assess the quantity and purity of the RNA isolates, spectrophotometrically, readings of each total-RNA were taken at 260 and 280 nm. All RNA sample A°260/A°280 ratios fell between 1.75 and 1.9, indicating very pure RNA in all cases. The quality of each RNA sample has been assessed by denaturing agarose gel electrophoresis and ethidium bromide staining. Two sharp bands (28-S and 18-S ribosomal RNA bands) segregated at the proper size (Grubor *et al*., 2004). First Strand cDNA synthesis has been conducted using READY TO GO YOU prime-First Strand Beads kit (Amersham).

PCR reactions: A PCR cocktail consisting of 1.0 μM upper and lower primers and 0.2 mM dNTPs, 10 mM Tris (pH 9.0), 50 mM KCL, 1.5 mM $MgCl_2$, 0.01% gelatin (W/V), 0.1 triton X-100 and 1.25 unit of taq polymerase was aliquoted into tubes with 100 ng DNA or cDNA. The samples were subjected to 35 cycles of PCR amplification. Each cycle includes denaturation at 95°C for 1 min, annealing at 58°C for 2 min and primer polymerization at 72°C for 1 min. The SOD1 primers were 5-GTG CTG AAG GGC GAC G-3 for sense and 5-TTT CCA CCT TTG CCC AAG-3 for antisense (Lyons *et al*., 1997).

PCR products were subjected to 1.5% agarose gel electrophoresis and stained with ethidium bromide. The gels were examined with a UV lamp at a wave length of 312 nm and were photographed using Mp4 plus Polaroid Camera.

The internal sequencing of SOD1 from both genomic DNA and cDNA has been performed using ABI PRISM version 3.7. Sequence alignment was carried out using NCBI-BlASTN 2.2.13 version (Altschul *et al*., 1997).

RESULTS

In the present investigation, a pair of primers specific for SOD1 used to test for the presence of this gene on total RNA and genomic DNA extracted from blood of native cattle in Egypt, using the PCR and Reverse Transcription (RT-PCR) techniques. A single sharp band at approximately 370 bp is obtained with both cDNA and genomic DNA (Fig. 1).

NCBI-Blast analysis of the internal sequence of SOD1 amplified PCR product (amplicon) of genomic DNA (Fig. 2) and of SOD1 amplified cDNA (Fig. 3) showed a 97% alignment (Fig. 4). The internal sequence analysis of SOD1 amplicon obtained from the cattle genomic DNA

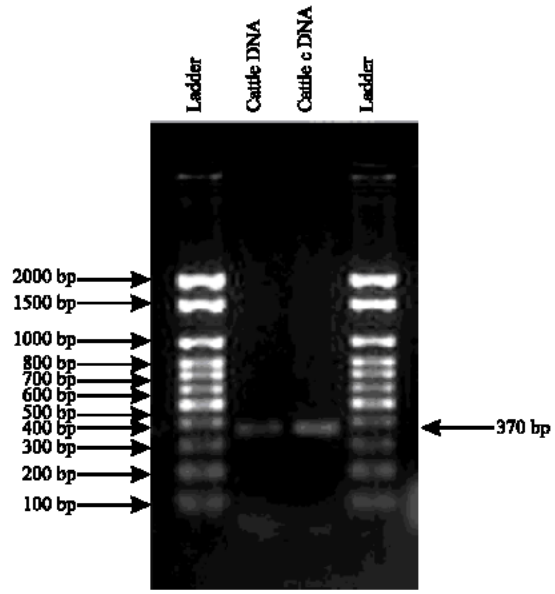

Fig. 1: Representive ethidium bromide-staind gel of amplified PCR products of SOD1 using cDNA and DNA extracted from blood Egyptian native cattle, L 100 bp and the left arrow indicate the amplified fragments size

showed a 98% alignment to mRNA of Bos taurus SOD1 sequences of the Gene Bank database (gene bank accession numbers, gi: 31341527, 7356542 and 162960) Fig. 5, represents the alignment of DNA with mRNA of Bos taurus SOD1 gi: 31341527 as an example). The obtained amplicon from the cattle genomic DNA also shares sequence homology with mRNA from other species such as Homo sapiens gi: 48762945 (87%), *Capra hircus* gi: 5865327 (95%) and *Mucaca mulatto* gi: 74136166 (91%). Figure 6 represents the alignment with *Homo sapiens*.

The amplicon sequence of the native cattle genomic DNA from nucleotide (nt) 15 to 306 aligns with four separate segments of Bos taurus chromosome 1 genomic contig ref |NW-929793.1| superoxide dismutase 1 (Fig. 7). Eighty one nucleotides (from 15 to 95 nt) showed 96% alignment with a segment of exon 4 of the Bos taurus chromosome 1 from nt coordinate 10406 to 10326 (Fig. 7A); 73 nucleotides (93 to 165) showed 98% sequence homology with exon 3 from nt coordinate 9501 to 9429 (Fig. 7B); 96 nucleotides (163 to 258) showed 98% sequence homology with a segment of exon 2 from nt coordinate 7649 to 7554 (Fig. 7C) and 53 nucleotides (254 to 306) showed 98% sequence homology with a segment of Bos taurus chromosome 1 from nt coordinate 3959 to 3908 (Fig. 7D). It also shares sequence homology (99%) with Bos taurus chromosome BTA13 genomic contig ref |NW-928647.1| similar to Superoxide dismutase (Fig. 8).

Moreover, blast analysis of the SOD1 amplicon sequence (370 nt) of the native cattle genomic DNA aligns with three separate segments of DNA on human chromosome 21, gi: 7768676 (Fig. 9) with sequence identities of 86, 94 and 83%. These three separate segments represent exon four (Fig. 9A); exon three (Fig. 9B) and exon two (Fig. 9C), respectively. In addition four interesting alignments are found; alignment of the amplicon of the cattle genomic DNA fragment showed 78% identities (Fig. 10) with human SOD1 processed pseudogene Ψ 69.1 (gi: 181206); 84% alignment (Fig. 11) with processed pseudogene Ψ 71.4 (gi: 180712); 80% alignment (Fig. 12) with human chromosome 16 DNA from nt coordinate 94952 to 95227 (gi: 29124036) and 84% alignment with chromosome 8 (gi: 21307426) from nt coordinate 70966 to 71135 (Fig. 13).

Study of the alignment of the two processed pseudogenes with human chromosomes 8 and 16 reveals that processed pseudogene Ψ 69.1 (gi: 181206) is 94% identical with chromosome 16 and only 80% with chromosome 8. Whereas the processed pseudogene Ψ 71.4 (gi: 180712) is 98% identical with chromosome 8 and 80% alignment with chromosome 16.

A series of alignment was carried out in order to identify the complete sequence of the SOD1 pseudogenes 69.1 and Ψ 71.4 on human chromosome 16 and 8, respectively. A segment of 2000 bases of chromosome 16 (from nt coordinate 94000 to 96000 which comprises the segment of human chromosome 16 (from nt coordinate 94952 to 95227) that gave an 80% alignment with the investigated amplicon is tested for alignment with the human SOD1 processed pseudogenes Ψ 69.1 (gi: 181206). Pair-wise blast analysis showed that a segment of 713 bases (from nt coordinate 94713 to 95423 including 2 gaps at 94803 and 95040) has 94% homology and a small fragment of 36 nt (from 94647 to 94682) showed 100% alignment. The 30 nt inbetween (from nt coordinate 94683 to 94712) failed to give significant alignment with Ψ 69.1 SOD1 processed pseudogene (Fig. 14). This segment found to be a poly A sequence (Fig. 16).

The whole sequence of retropseudogene on chromosome 16 has high identities with exon 2, 3, 4 (82, 86 and 82%, respectively) and partially with exon 5 (88%) and no alignment with exon 1 of SOD1 normal gene on human chromosome 21 (Fig. 15). The complete sequence (777 nt) of SOD1 retropseudogene is represented in Fig. 16.

Similarly, NCBI-Blast analysis of a 2000 bases (from 70000 to 72000) that comprises the human segment of chromosome 8 (from 70966 to 71135) which gave 84% alignment with the investigated amplicon is tested for alignment with processed pseudogene Ψ 71.4 (gi: 180712). The alignment shows that a segment of 668 bases

AACGACCTGANTCNTGAGATCAGAGAATCTACAATATCCACGACGGCAACACCGTTTTTG
TCAGCTGTCACATTGCCCAGGTCTCCAACATGCCTCTCTTCATCTTTTGGCCCACTGTGTTT
TTTGGACAGAGGATTAAAGTGAGGACCTGCACTGGTACAGCCTTGTGTATTGTCTCCAAAC
TGATGGACGTGGAATCCATGATCACCTTCAGTCAATCCTGTAATGGATCCAGTTACC
ACGACTGTATTTCCCTTTGCCTCGAAGTGGATGGTGCCTTGCACCGGGCCGTCGCCCCTTC
AGCACAAAGANNNNNTTCNTCGNCCTTNAGCACANA

Fig. 2: Nucleotide sequences of the reverse strand superoxide dismutase-1 (SOD1) for the amplified PCR product of DNA extracted from blood Egyptian native cattle

TTCTCCTGAGAGTGAAATCAGAGGATCTACAATATCCACGANGGCAACACCGTTTTTGTCA
GCTGTCACATTGCCCAGGTCTCCAACATGCCTCTCTTCATCTTTTGGCCCACCGTGTTTTTT
GGACAGAGGATTAAAGTGAGGACCTGCACTGGTACAGCCTTGTGTATTGTCTCCAAACTG
ATGGACGTGGAATCCATGATCACCTTCAGTCAATCCTGTAATGGATCCAGTTACCAC
GACTGTATCTCCCTTTGCCTCGAAGTGGATGGTGCCTTGCACGGGGCCGCNCCCCTTTNCA

Fig. 3: Nucleotide sequences of the reverse strand superoxide dismutase-1 (SOD1) for the amplified RT-PCR product of mRNA extracted from blood Egyptian native cattle

```
a)     : 13  tgaaatcagaggatctacaatatccacganggcaacaccgtttttgtcagctgtcacatt 72
           ||| ||||||| |||||||||||||||| |||||||||||||||||||||||||||||||
(b)    : 15  tgagatcagagaatctacaatatccacgacggcaacaccgtttttgtcagctgtcacatt 74

(a)    : 73  gcccaggtctccaacatgcctctcttcatcttttggcccaccgtgtttttttggacagagg 132
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
(b)    : 75  gcccaggtctccaacatgcctctcttcatcttttggcccactgtgtttttttggacagagg 134

(a)    : 133 attaaagtgaggacctgcactggtacagccttgtgtattgtctccaaactgatggacgtg 192
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
(b)    : 135 attaaagtgaggacctgcactggtacagccttgtgtattgtctccaaactgatggacgtg 194

(a)    : 193 gaatccatgatcaccttcagtcaatcctgtaatggatccagttaccacgactgtatctcc 252
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
(b)    : 195 gaatccatgatcaccttcagtcaatcctgtaatggatccagttaccacgactgtatttcc 254

(a)    : 253 ctttgcctcgaagtggatggtgccttgcacggggccg 289
             |||||||||||||||||||||||||||||| ||||||
(b)    : 255 ctttgcctcgaagtggatggtgccttgcaccgggccg 291
```

Fig. 4: Alignment of the SOD1 amplified fragment of cDNA (a) with the SOD1 amplified fragment of DNA (b) extracted from Egyptian native cattle, Identities = 271/277 (97%) Strand = Plus/Plus

```
(a)    19   ATCAGAGAATCTACAATATCCACGACGGCAACACCGTTTTTGTCAGCTGTCACATTGCCC  78
            ||||||| |||||||||||||||| |||||||||||||||||||||||||||||||||||
(b)    387  ATCAGAGGATCTACAATATCCACGATGGCAACACCGTTTTTGTCAGCTGTCACATTGCCC  328

(a)    79   AGGTCTCCAACATGCCTCTCTTCATCTTTTGGCCCACTGTGTTTTTTGGACAGAGGATTA  138
            |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
(b)    327  AGGTCTCCAACATGCCTCTCTTCATCTTTTGGCCCACCGTGTTTTTTGGACAGAGGATTA  268

(a)    139  AAGTGAGGACCTGCACTGGTACAGCCTTGTGTATTGTCTCCAAACTGATGGACGTGGAAT  198
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
(b)    267  AAGTGAGGACCTGCACTGGTACAGCCTTGTGTATTGTCTCCAAACTGATGGACGTGGAAT  208

(a)    199  CCATGATCACCTTCAGTCAATCCTGTAATGGATCCAGTTACCACGACTGTATTTCCCTTT  258
            |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
(b)    207  CCATGATCACCTTCAGTCAATCCTGTAATGGATCCAGTTACCACGACTGTATCTCCCTTT  148

(a)    259  GCCTCGAAGTGGATGGTGCCTTGCACCGGGCCGTCGCCCCTTCAGCAC  306
            |||||||||||||||||||||||||||||||||||| ||||||||||
(b)    147  GCCTCGAAGTGGATGGTGCCTTGCACCGGGCCGTCG-CCCTTCAGCAC  10
```

Fig. 5: Alignment of the SOD1 amplified fragment of Egyptian native cattle DNA (a), with Bos taurus superoxide dismutase 1, soluble (SOD1), mRNA (b) gi: 31341527, Identities = 283/288 (98%), Gaps = 1/288 (0%) Strand = Plus/Minus

```
(a)       15   TGAGATCAGAGAATCTACAATATCCACGACGGCAACACCGTTTTTGTCAGCTGTCACATT  74
               ||||||||| ||||||| |||||  |||  |||| ||||| | ||||||||| ||||||||
(b)       466  TGAGATCACAGAATCTTCAATAGACACATCGGCCACACCATCTTTGTCAGCAGTCACATT  407

(a)       75   GCCCAGGTCTCCAACATGCCTCTCTTCATCTTTTGGCCCACTGTGTTTTTTGGACAGAGG  134
               |||||| ||||||||||||||||||||||| |||||||||| ||||||| |||| |||||
(b)       406  GCCCAAGTCTCCAACATGCCTCTCTTCATCCTTTGGCCCACCGTGTTTTCTGGATAGAGG  347

(a)       135  ATTAAAGTGAGGACCTGCACTGGTACAGCCTTGTGTATTGTCTCCAAACTGATGGACGTG  194
               |||||||||||||||||||||||||||||  |||||| |||||||||| ||| || ||
(b)       346  ATTAAAGTGAGGACCTGCACTGGTACAGCCTGCTGTATTATCTCCAAACTCATGAACATG  287

(a)       195  GAATCCATGATCACCTTCAGTCAATCCTGTAATG  228
               |||||||||    |||||||||| |||| |||||
(b)       286  GAATCCATGCAGGCCTTCAGTCAGTCCTTTAATG  253
```

Fig. 6: Alignment of the SOD1 amplified fragment of Egyptian native cattle DNA (a) with Homo sapiens (b) superoxide dismutase 1, soluble (amyotrophic lateral sclerosis 1 (adult) (SOD1), mRNA gi: 48762945, Identities = 187/214 (87%), Gaps = 0/214 (0%) Strand = Plus/Minus

```
A
Identities = 78/81 (96%), Gaps = 0/81 (0%)
 Strand=Plus/Minus
DNA cattle    1  TGAGATCAGAGAATCTACAATATCCACGACGGCAACACCGTTTTTGTCAGCTGTCACATT  74
                 ||| ||||||| |||||||||||||||| |||||||||||||||||||||||||||||
BTA1      10406  TGAAATCAGAGGATCTACAATATCCACGATGGCAACACCGTTTTTGTCAGCTGTCACATT  10347

DNA cattle    75     GCCCAGGTCTCCAACATGCCT  95
                     |||||||||||||||||||||
BTA1          10346  GCCCAGGTCTCCAACATGCCT  10326
B
Identities = 72/73 (98%), Gaps = 0/73 (0%)
 Strand=Plus/Minus
DNA cattle    93    CCTCTCTTCATCTTTTGGCCCACTGTGTTTTTTGGACAGAGGATTAAAGTGAGGACCTGC  152
                    ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
BTA1          9501  CCTCTCTTCATCTTTTGGCCCACCGTGTTTTTTGGACAGAGGATTAAAGTGAGGACCTGC  9442

DNA cattle    153   ACTGGTACAGCCT  165
                    |||||||||||||
BTA1          9441  ACTGGTACAGCCT  9429
C
Identities = 95/96 (98%), Gaps = 0/96 (0%)
 Strand=Plus/Minus
DNA cattle    163   CCTTGTGTATTGTCTCCAAACTGATGGACGTGGAATCCATGATCACCTTCAGTCAATCCT  222
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
BTA1          7649  CCTTGTGTATTGTCTCCAAACTGATGGACGTGGAATCCATGATCACCTTCAGTCAATCCT  7590

DNA cattle    223   GTAATGGATCCAGTTACCACGACTGTATTTCCCTTT  258
                    |||||||||||||||||||||||||||| |||||||
BTA1          7589  GTAATGGATCCAGTTACCACGACTGTATCTCCCTTT  7554
D
Identities = 52/53 (98%), Gaps = 1/53 (1%)
 Strand=Plus/Minus
DNA cattle    254   CCTTTGCCTCGAAGTGGATGGTGCCTTGCACCGGGCCGTCGCCCCTTCAGCAC  306
                    |||||||||||||||||||||||||||||||||||||||| ||||||||||
BTA1          3959  CCTTTGCCTCGAAGTGGATGGTGCCTTGCACCGGGCCGTCG-CCCTTCAGCAC  3908
```

Fig. 7: Alignment of the SOD1 amplified fragment of DNA cattle with Bos taurus chromosome 1 genomic contigref|NW_929793.1|Bt1_WGA5_2 superoxide dismutase 1, soluble Length = 691109

```
Strand=Plus/Plus

(a)    15      TGAGATCAGAGAATCTACAATATCCACGACGGCAACACCGTTTTTGTCAGCTGTCACATT  74
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
(b)    171375  TGAGATCAGAGAATCTACAATATCCACGACGGCAACACCGTTTTTGTCAGCTGTCACATT  171434

(a)    75      GCCCAGGTCTCCAACATGCCTCTCTTCATCTTTTGGCCCACTGTGTTTTTTGGACAGAGG  134
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
(b)    171435  GCCCAGGTCTCCAACATGCCTCTCTTCATCTTTTGGCCCACTGTGTTTTTTGGACAGAGG  171494

(a)    135     ATTAAAGTGAGGACCTGCACTGGTACAGCCTTGTGTATTGTCTCCAAACTGATGGACGTG  194
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
(b)    171495  ATTAAAGTGAGGACCTGCACTGGTACAGCCTTGTGTATTGTCTCCAAACTGATGGACGTG  171554

(a)    195     GAATCCATGATCACCTTCAGTCAATCCTGTAATGGATCCAGTTACCACGACTGTATTTCC  254
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
(b)    171555  GAATCCATGATCACCTTCAGTCAATCCTGTAATGGATCCAGTTACCACGACTGTATTTCC  171614

(a)    255      CTTTGCCTCGAAGTGGATGGTGCCTTGCACCGGGCCGTCGCCCCTTCAGCAC  306
                ||||||||||||||||||||||||||||||||||||||||||| ||||||||
(b)     171615  CTTTGCCTCGAAGTGGATGGTGCCTTGCACCGGGCCGTCGCCC-TTCAGCAC  171665
```

Fig. 8: Alignment of the SOD1 amplified fragment of DNA cattle (a) with Bos taurus chromosome 13 (b) genomic contig ref |NW_928647.1|Bt13_WGA2317_2 similar to Superoxide dismutase Identities = 291/292 (99%), Gaps = 1/292 (0%)

```
(A, part of Exon-4)
  Identities = 70/81 (86%)

  Human: 310226    aggcatgttggagacttgggcaatgtgactgctgacaaagatggtgtggccgatgtgtct 310285
                   ||||||||||||||| |||||||||||||| ||||||||| | ||||| ||||  |||  |
  Cattle: 95       aggcatgttggagacctgggcaatgtgacagctgacaaaaacggtgttgccgtcgtggat 36

  Human : 310286    attgaagattctgtgatctca 310306
                    |||| ||||||| ||||||||
  Cattle:  35       attgtagattctctgatctca 15

 (B, Exon -3)
 Identities = 69/73 (94%)

 Human: 309417    aggctgtaccagtgcaggtcctcactttaatcctctatccagaaaacacggtgggccaaa 309476
                  |||||||||||||||||||||||||||||||||||| |||| ||||||| ||||||||||
 Cattle :   165   aggctgtaccagtgcaggtcctcactttaatcctctgtccaaaaaacacagtgggccaaa 106

 Human :   309477 ggatgaagagagg 309489
                  |||||||||||||
 Cattle :  105    agatgaagagagg 93

(C, part of Exon-2)
 Identities = 51/61 (83%)

Human: 306792    cattaaaggactgactgaaggcctgcatggattccatgttcatgagtttggagataatac 306851
                  ||||| |||| |||||||||||    |||||||||||| || ||| |||||||||| |||||
Cattle: 228      cattacaggattgactgaaggtgatcatggattccacgtccatcagtttggagacaatac 169

Human:  306852   a 306852
                 |
Cattle: 168      a 168
```

Fig. 9: Alignment of the SOD1 amplified fragment of Egyptian native cattle DNA with human chromosome 21 gi: 7768676, Strand = Plus/Minus

```
 DNA cattle        15    tgagatcagagaatctacaatatccacgacggcaacaccgtttttgtcagctgtcacatt 74
                         ||||| |||||||||| ||| |  || |  ||| ||||  | |||| |||| | ||||||
 pseudogene        458   tgagaccagagaatcttcaacagacatgttggcgacaca-tctttgccagcggccacatt 400

 DNA cattle:       75    gcccaggtctccaacatgcctctcttcatcttttggcccactgtgtttttttggacagagg 134
                         || |||||||||| |||||||||| |||  |||| |||| ||| |||||||  | |||
 pseudogene        399   tcc-aggtctccaacctgcctctcttgatcccttggtccaccgtg-tttttggttaaagg 342

 DNA cattle:       135   attaaagtgaggacctgcactggtacagccttgtgtattgtctccaaactgatggacgtg 194
                         |||||||| ||||  ||| || |||||||| ||||| |   |||||||||||||| || ||
  pseudogene       341   attaaagtaagga-atgccct-gtacagcc-tgtgt-tatactccaaactgatgaacatg 286

 DNA cattle    :   195    gaatccatgatcaccttcagtcaatcctgtaatg 228
                         |||||  || |  | ||||||||||||||||||||
 pseudogene        285   gaatctgtgctggcgttcagtcaatcctgtaatg 252
```

Fig. 10: Alignment of the SOD1 amplified fragment of Egyptian native cattle DNA with Human Cu/Zn processed pseudogene Ψ 69.1 gi: 181206, Identities = 169/214 (78%), Gaps = 7/214 (3%). Strand = Plus/Minus

```
Query  57   TTTGTCAGCTGTCACATTGCCCAGGTCTCCAACATGCCTCTCTTCATCTTTTGGCCCACT  116
            |||| |||| |||||||||||||| | |  |||||||||||||| ||| ||||| ||||
Sbjct  377  TTTGCCAGCAGTCACATTGCCCAG-TGTTTAACATGCCTCTCTTGATCCTTTGGTCCACC  319

Query  117  GTGTTTTTTGGACAGAGGATTAAAGTGAGGACCTGCACTGGTACAGCCTTGTGTATTGTC  176
            |||||||  ||| |||||||||||||||||||||||||||||||||||||||| |||| |
Sbjct  318  GTGTTTTGGGGATAGAGGATTAAAGTGAGGACCTGCACTGGTACAGCCTTGTATATTATT  259

Query  177  TCCAAACTGATGGACGTGGAATCCATGATCACCTTCAGTCAATCCTGTAATGGATCC  233
            | || ||||||| || ||| |||  || |  ||||||| |||||||||||||| |||
Sbjct  258  TGCACACTGATGAACATGGCATC--TGCTGGCCTTCAGGCAATCCTGTAATGGTTCC  204
```

Fig. 11: Alignment of the SOD1 amplified fragment of Egyptian native cattle DNA with Human Cu/Zn processed pseudogene ψ 71.4 gi:180712, Identities = 149/177 (84%), Gaps = 3/177 (1%) Strand = Plus/Minus

```
Human: 94952 tgagaccagagaatcttcaacagacatgttggcgacaca-tctttgccagcggccacatt 95010
             ||||| |||||||||| ||| |  || |  ||| ||||  | |||| |||| | ||||||
cattle 15    tgagatcagagaatctacaatatccacgacggcaacaccgtttttgtcagctgtcacatt 74

Human: 95011 acccaggtctccaacctgcctctcttgatcctt-ggcccactgtggtttttggttaaagg 95069
             |||||||||||||| |||||||||| ||| || |||||||||| |||||||  | |||
cattle: 75   gcccaggtctccaacatgcctctcttcatcttttggcccactgtgttttttggacagagg 134

Human: 95070 attaaagtaaggaactgccctggtacagcctggtgtattatctccaaactgatgaacatg 95129
             |||||||| |||| |||| |||||||||||| ||||||| |||||||||||||| || ||
cattle: 135  attaaagtgaggacctgcactggtacagccttgtgtattgtctccaaactgatggacgtg 194

Human: 95130 gaatctgtgctggcgttcagtcaatcctgtaatgcattctgacaccataaatggttcatt 95189
             |||||  || |  | ||||||||||||||||||| || | |  ||||  | ||    |||
cattle: 195  gaatccatgatcaccttcagtcaatcctgtaatggatccagttaccacgactg---tatt 251

Human: 95190 ttccttctgctcaaagtggatggtgccctgcaccgggc 95227
             | ||||   ||| |||||||||||||| |||||||||
cattle: 252  tccctttgcctcgaagtggatggtgccttgcaccgggc 289
```

Fig. 12: Alignment of the SOD1 amplified fragment of Egyptian native cattle DNA with human chromosome 16 gi: 29124036, Identities = 223/278(80%), Gaps = 5 /278 (1%). Strand = Plus/Plus

```
Query: 70966 tttgccagcagtcacattgcccaggtgtttaacatgcctctcttgatcctttggtccacc 71025
              |||| |||| |||||||||||||||| |  |||||||||||||| ||| ||||| ||||
Sbjct: 57    tttgtcagctgtcacattgcccaggtctccaacatgcctctcttcatcttttggcccact 116


Query: 71026 gtgttttggggatagaggattaaagtgaggacctgcactggtacagccttgtatattatt 71085
              |||||||  ||| |||||||||||||||||||||||||||||||||||||||| |||| |
Sbjct: 117   gtgtttttttggacagaggattaaagtgaggacctgcactggtacagccttgtgtattgtc 176


Query: 71086 tgcacactgatgaacatggcatc--tgctggccttcaggcaatcctgtaatg 71135
              | || ||||||| || ||| |||  || |  ||||||| |||||||||||||
Sbjct: 177   tccaaactgatggacgtggaatccatgatcaccttcagtcaatcctgtaatg 228
```

Fig. 13: Alignment of the SOD1 amplified fragment of Egyptian native cattle DNA with Human chromosome 8 gi:21307426 Identities = 146/172 (84%), Gaps = 2/172 (1%) Strand = Plus/Plus

```
Identities = 36/36 (100%)

Ch. 16      94647  CACTCCAACCTGGGCTACAGAGTGAGACTCTGCCAC  94682
                   ||||||||||||||||||||||||||||||||||||
pseudogene  759    CACTCCAACCTGGGCTACAGAGTGAGACTCTGCCAC  724


Identities = 673/713 (94%), Gaps = 22/713 (3%)
Strand=Plus/Minus

Ch. 16   94713     GATTTCAGTGTTTAATGTTTATCAAATTTCATCTCTACAGCTAGCAGGATAATAGATGAG  94772
                   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||
pseudogene  693    GATTTCAGTGTTTAATGTTTATCAAATTTCATCTCTACAGCTAGCAGGATAAC-GATGAG  635

Ch. 16      94773  TTAATGGAACTCAGACCACATTTAAGAGAA-TGTTTATTGGGCAATCCCAATTACACCAC  94831
                   ||||  | | | ||| |||||| ||||||| |||||||||| ||| ||||||||||||||
pseudogene  634    TTAAGAGCA-TGAGAGCACATT-AAGAGAAATGTTTATTGG-CAA-CCCAATTACACCAC  579

Ch. 16      94832  AAACCAAACAACTTCTAGCATTTCCTGTGTTTGTACTTTCTTCATTTTCACCTTTGCCCA  94891
                   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
pseudogene  578    AAACCAAACAACTTCTAGCATTTCCTGTGTTTGTACTTTCTTCATTTTCACCTTTGCCCA  519

Ch. 16      94892  AGTCATCTGGTTTGTCATGGACACCATTGTGTGGGCAGTGATGGAATAGTGTCCTGAGAG  94951
                   |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
pseudogene  518    AGTCATCTGGTTTGTCATGGACACCATTGTGTGGGCAGTGATGGAATAGTCTCCTGAGAG  459

Ch. 16      94952  TGAGACCAGAGAATCTTCAACAGACATGTTGGCGACACATCTTTGCCAGCGGCCACATTA  95011
                   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
pseudogene  458    TGAGACCAGAGAATCTTCAACAGACATGTTGGCGACACATCTTTGCCAGCGGCCACATT-  400

Ch. 16     95012   CCCAGGTCTCCAACCTGCCTCTCTTGAT-CCTTGGCCCACTGTGGTTTTTGGTTAAAGGA  95070
                     |||||||||||||||||||||||||| |||||| |||| || ||||||||||||||||
pseudogene  399    TCCAGGTCTCCAACCTGCCTCTCTTGATCCCTTGGTCCACCGT-GTTTTTGGTTAAAGGA  341

Ch. 16      95071  TTAAAGTAAGGAACTGCCCTGGTACAGCCTGGTGTATTATCTCCAAACTGATGAACATGG  95130
                   |||||||||||||| |||||| ||||||||| ||||  || ||||||||||||||||||||
pseudogene  340    TTAAAGTAAGGAA-TGCCCT-GTACAGCCT-GTGTTATA-CTCCAAACTGATGAACATGG  285

Ch. 16      95131  AATCTGTGCTGGCGTTCAGTCAATCCTGTAATGCATTCTGACACCATAAATGGTTCATTT  95190
                   |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
pseudogene  284    AATCTGTGCTGGCGTTCAGTCAATCCTGTAATGC-TTCTGACACCATAAATGGTTCATTT  226

Ch. 16      95191  TCCTTCTGCTCAAAGTGGATGGTGCCCTGCACCGGGCTCTGGCCCTTCAACATGTACATG  95250
                   |||||||||||||| ||||| ||  ||||||||||||||||||||||||||||||||||||
pseudogene  225    TCCTTCTGCTCAAA-TGGAT-GTCGCCTGCACCGGGCTCTGGCCCTTCAACATGTACATG  168

Ch. 16      95251  ACCTTCATCATTATGACTCCAGGGGCCACACTCTGGTCTGGGCTCTGAGAGGTGTGAGGG  95310
                   |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
pseudogene  167    ACCTTCATCATTATGACTCCAGGGGCCACACTCT-GTCTGGGCTCTGAGAGGTGTGAGGG  109

Ch. 16      95311  ACACAGGAGACACTGCAGGAGATGGCAGCACATGCTGACTTTCAAATGATGAGCCTGTAA  95370
                   ||||||||||||||||  || |||||||||||| ||||| ||||||||||||||||||||
pseudogene  108    ACACAGGAGACACTGCCAGA-ATGGCAGCACATACTGAC-TTCAAATGATGAGCCTGTAA  51

Ch. 16      95371  CACAGATCACTTTGGCTTCAATATAGGAACAGGTAGGTAAGGAGAGGGGCTGC  95423
                   ||||||||||||||   |||||||| ||||||||||||||||||||||||||
pseudogene  50     CACAGATCACTTTG--CTCAATATA-GAACAGGTAGGTAAGGAGAGGGGCTGC  1
```

Fig. 14: Alignment of human chromosome 16 gi: 29124036 from nt coordinate 94000 to 96000 with Human Cu/Zn processed pseudogene ψ 69.1 gi: 181206

```
(Exon-5)
Identities = 178/201 (88%), Gaps = 0/201 (0%)
 Strand=Plus/Minus
retropseudogene 67       GATTTCAGTGTTTAATGTTTATCAAATTTCATCTCTACAGCTAGCAGGATAATAGATGAG  126
                         |||| ||||||||||||||||||||    | ||| |||||||||||||||||||| |||||||
chromosome 21  311641    GATTACAGTGTTTAATGTTTATCAGGATACATTTCTACAGCTAGCAGGATAACAGATGAG  311582

retropseudogene  127     TTAATGGAACTCAGACCACATTTAAGAGAATGTTTATTGGGCAATCCCAATTACACCACA  186
                         |||| ||  ||||||| ||||  ||| ||||||||||||||| ||||||||||||||||
chromosome 21   311581   TTAAGGGGCCTCAGACTACATCCAAGGGAATGTTTATTGGGCGATCCCAATTACACCACA  311522

retropseudogene 187      AACCAAACAACTTCTAGCATTTCCTGTGTTTGTACTTTCTTCATTTTCACCTTTGCCCAA  246
                         | |||||| ||||| ||| |||||||| |||||||||||||||||| ||||||||||||
chromosome 21  311521    AGCCAAACGACTTCCAGCGTTTCCTGTCTTTGTACTTTCTTCATTTCCACCTTTGCCCAA  311462

retropseudogene 247      GTCATCTGGTTTGTCATGGAC  267
                         ||||||||| ||| ||||||||
chromosome 21 311461     GTCATCTGCTTTTTCATGGAC  311441


(Exon-4)
Identities = 99/120 (82%), Gaps = 1/120 (0%)
 Strand=Plus/Minus
retropseudogene  267      CACCATTGTGTGGGCAGTGATGGAATAGTGTCCTGAGAGTGAGACCAGAGAATCTTCAAC  326
                         ||||| |||| || || ||||| ||| || ||||||||||||||| || ||||||||||
chromosome 21   310345   CACCAGTGTGCGGCCAATGATGCAATGGTCTCCTGAGAGTGAGATCACAGAATCTTCAAT  310286

retropseudogene 327      AGACATGTTGGCGACA-CATCTTTGCCAGCGGCCACATTACCCAGGTCTCCAACCTGCCT  385
                         |||||  | ||| ||| |||||||| |||| | |||||| |||| ||||||||| |||||
chromosome 21  310285    AGACACATCGGCCACACCATCTTTGTCAGCAGTCACATTGCCCAAGTCTCCAACATGCCT  310226


(Exon-3)
Identities = 63/73 (86%), Gaps = 1/73 (1%)
 Strand=Plus/Minus
retropseudogene 383       CCTCTCTTGATCCTT-GGCCCACTGTGGTTTTTGGTTAAAGGATTAAAGTAAGGAACTGC  441
                          ||||||| |||||| ||||||| ||| ||| ||| || |||||||||| |||| ||||
chromosome 21  309489     CCTCTCTTCATCCTTTGGCCCACCGTGTTTTCTGGATAGAGGATTAAAGTGAGGACCTGC  309430

retropseudogene 442      CCTGGTACAGCCT  454
                         |||||||||||
chromosome 21  309429    ACTGGTACAGCCT  309417


(Exon-2)
Identities = 76/92 (82%), Gaps = 0/92 (0%)
 Strand=Plus/Minus
Retropseudogene  452     CCTGGTGTATTATCTCCAAACTGATGAACATGGAATCTGTGCTGGCGTTCAGTCAATCCT  511
                               |||| |||||||||||||||| ||||||||||||||  ||| ||| |||||||| ||||
chromosome 21   306857  CCTGCTGTATTATCTCCAAACTCATGAACATGGAATCCATGCAGGCCTTCAGTCAGTCCT  306798

retropseudogene  512      GTAATGCATTCTGACACCATAAATGGTTCATT  543
                         |||||| | |  ||||| | | |||| ||||
chromosome 21   306797   TTAATGCTTCCCCACACCTTCACTGGTCCATT  306766
```

Fig. 15: Alignment of SOD1 retropseudogene on human chromosome 16 with complete sequence of human chromosome 21 gi: 7768676

```
94621 tgcagtgagc cgagatcatg ccactgcact ccaacctggg ctacagagtg agact ctgcc
94681 acaaaaaaaa ataaaataaa ataaaaaata aagatttcag tgtttaatgt ttatcaaatt
94741 tcatctctac agctagcagg ataatagatg agttaatgga actcagacca catttaagag
94801 aatgtttatt gggcaatccc aattacacca caaaccaaac aacttctagc atttcctgtg
94861 tttgtacttt cttcattttc acctttgccc aagtcatctg gtttgtcatg gacaccattg
94921 tgtgggcagt gatggaatag tgtcctgaga gtgagaccag agaatcttca acagacatgt
94981 tggcgacaca tctttgccag cggccacatt acccaggtct ccaacctgcc tctcttgatc
95041 cttggcccac tgtggttttt ggttaaagga ttaaagtaag gaactgcctt ggtacagcct
95101 ggtgtattat ctccaaactg atgaacatgg aatctgtgct ggcgttcagt caatcctgta
95161 atgcattctg acaccataaa tggttcattt tccttctgct caaagtggat ggtgccctgc
95221 accgggctct ggccccttcaa catgtacatg accttcatca ttatgactcc aggggccaca
95281 ctctggtctg ggctctgaga ggtgtgaggg acacaggaga ca ctgca gga gatggcagca
95341 catgctgact ttcaaatgat gagcctgtaa cacagatcac tttggcttc a atat aggaac
95401 aggtaggtaa ggagaggggc tgc
```

Fig. 16: The full sequence of SOD1 retropseudogene on human chromosome 16 gi: 29124036 from nt coordinate 94647 to 95423

```
Score = 61.9 bits (31),  Expect = 7e-06
 Identities = 31/31 (100%), Gaps = 0/31 (0%)
 Strand=Plus/Minus


Query  70626  TAAGAAGTTTCTGGAGAATTGTGAATGTGTG  70656
              ||||||||||||||||||||||||||||||
Sbjct  716    TAAGAAGTTTCTGGAGAATTGTGAATGTGTG  686


Identities = 658/668 (98%), Gaps = 4/668 (0%)
 Strand=Plus/Minus


Query  70671  CTTTACACTTTTAAGATTACAGTGTTAAATGTTTATCAAGTTACATTTCTACAGCTAGCA  70730
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  671    CTTTACACTTTTAAGATTACAGTGTTAAATGTTTATCAAGTTACATTTCTACAGCTAGCA  612

Query  70731  GGATAACACATGTGTTACTGGCACTCAGACCTCATCCTAGGGAATGTTTACTGGGAATTC  70790
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  611    GGATAACACATGGGTTACTGGCACTCAGACCTCATCCTAGGGAATGTTTACTGGGAATTC  552

Query  70791  CCAAATACACCACAAGCCAAATAACTTCCAGCATTTCCCATCTTTGTACTTTCTTCATTT  70850
              ||||||||||||||| |||||||||||||||| |||||||||||||||||||||||||
Sbjct  551    CCAAATACACCACAAGTCAAATAACTTCCAGCATT-CCCATCTTTGTACTTTCTTCATTT  493

Query  70851  CCATCTTTGCCCAAGTCATCTGGTTTTTCATGGACCACCATTGCATGGCCAATGATGGAG  70910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  492    CCATCTTTGCCCAAGTCATCTGGTTTTTCATGGACCACCATTGCATGGCCAATGATGGAG  433

Query  70911  TGGCCTCCTGAGAGTGAGATCATAGAATCTTCAATAGACACATCATTGCATCATCTTTGC  70970
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  432    TGGCCTCCTGAGAGTGAGATCATAGAATCTTCAATAGACACATCATTGCATCATCTTTGC  373

Query  70971  CAGCAGTCACATTGCCCAGGTGTTTAACATGCCTCTCTTGATCCTTTGGTCCACCGTGTT  71030
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Sbjct  372    CAGCAGTCACATTGCCCAG-TGTTTAACATGCCTCTCTTGATCCTTTGGTCCACCGTGTT  314

Query  71031  TTGGGGATAGAGGATTAAAGTGAGGACCTGCACTGGTACAGCCTTGTATATTATTTGCAC  71090
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  313    TTGGGGATAGAGGATTAAAGTGAGGACCTGCACTGGTACAGCCTTGTATATTATTTGCAC  254

Query  71091  ACTGATGAACATGGCATCTGCTGGCCTTCAGGCAATCCTGTAATGTGTCCCAACACCACA  71150
              ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||
Sbjct  253    ACTGATGAACATGGCATCTGCTGGCCTTCAGGCAATCCTGTAATGGTTCCCAACACCACA  194

Query  71151  ACTGGTCCATTTTCCTTCTGCTTGAAATGGATAATTCTCGGTACCAGCCCTTCGCCCTTC  71210
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  193    ACTGGTCCATTTTCCTTCTGCTTGAAATGGATAATTCTCGGTACCAGCCCTTCGCCCTTC  134

Query  71211  AGCATGCACTGATGGGGCCACGCTCCTGTCTGGGTTCTGAGGATTGCCAGGGAAACAGGA  71270
              | |||||||||||||||||||||||||||||||||||||| |||||| ||||||||||||
Sbjct  133    AACATGCACTGATGGGGCCACGCTCCTGTCTGGGTTCTG-GGATTG-CAGGGAAACAGGA  76

Query  71271  GGTGCTATAAGAGATGACAACACAAACGCCAGAATTGTTAATGTTTTCACTGTCCATGAT  71330
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Sbjct  75     GGTGCTATAAGAGATGACAACACAAACCCCAGAATTGTTAATGTTTTCACTGTCCATGAT  16

Query  71331  GTATTTCT  71338
              ||||||||
Sbjct  15     GTATTTCT  8
```

Fig. 17: Allignment of human chromosome 8 from nt coordinate 70000 to 72000 gi:21307426 with Human Cu/Zn processed pseudogene 71.4 gi:180712

```
Exon 5
Identities = 190/214 (88%), Gaps = 0/214 (0%) Strand=Plus/Minus

Query  653     TTACACTTTTAAGATTACAGTGTTAAATGTTTATCAAGTTACATTTCTACAGCTAGCAGG  712
               ||||||||||||||||||||||||| |||||||||| | ||||||||||||||||||||
Sbjct  311653  TTACACTTTTAAGATTACAGTGTTTAATGTTTATCAGGATACATTTCTACAGCTAGCAGG  311594

Query  713     ATAACACATGTGTTACTGGCACTCAGACCTCATCCTAGGGAATGTTTACTGGGAATTCCC  772
               |||||| ||| ||||  ||  |||||||  ||||| |||||||||||| ||||   ||||
Sbjct  311593  ATAACAGATGAGTTAAGGGGCCTCAGACTACATCCAAGGGAATGTTTATTGGGCGATCCC  311534

Query  773     AAATACACCACAAGCCAAATAACTTCCAGCATTTCCCATCTTTGTACTTTCTTCATTTCC  832
               || ||||||||||||||  |||||||| |||||  ||||||||||||||||||||||||
Sbjct  311533  AATTACACCACAAGCCAAACGACTTCCAGCGTTTCCTGTCTTTGTACTTTCTTCATTTCC  311474

Query  833     ATCTTTGCCCAAGTCATCTGGTTTTTCATGGACC  866
               | ||||||||||||||||| ||||||||||||
Sbjct  311473  ACCTTTGCCCAAGTCATCTGCTTTTTCATGGACC  311440

Exon 4
Identities = 103/122 (84%), Gaps = 1/122 (0%) Strand=Plus/Minus

Query  864     ACCACCATTGCATGGCCAATGATGGAGTGGCCTCCTGAGAGTGAGATCATAGAATCTTCA  923
               ||||||| ||   |||||||||||| | ||| ||||||||||||||||| ||||||||||
Sbjct  310347  ACCACCAGTGTGCGGCCAATGATGCAATGGTCTCCTGAGAGTGAGATCACAGAATCTTCA  310288

Query  924     ATAGACACATC-ATTGCATCATCTTTGCCAGCAGTCACATTGCCCAGGTGTTTAACATGC  982
               ||||||||||      || |||||||| ||||||||||||||||| || |  |||||||
Sbjct  310287  ATAGACACATCGGCCACACCATCTTTGTCAGCAGTCACATTGCCCAAGTCTCCAACATGC  310228

Query  983     CT  984
               ||
Sbjct  310227  CT  310226

Exon 3
Identities = 69/73 (94%), Gaps = 0/73 (0%) Strand=Plus/Minus

Query  982     CCTCTCTTGATCCTTTGGTCCACCGTGTTTTGGGGATAGAGGATTAAAGTGAGGACCTGC  1041
               |||||||| ||||||||| ||||||||||||  ||||||||||||||||||||||||||
Sbjct  309489  CCTCTCTTCATCCTTTGGCCCACCGTGTTTTCTGGATAGAGGATTAAAGTGAGGACCTGC  309430

Query  1042    ACTGGTACAGCCT  1054
               |||||||||||||
Sbjct  309429  ACTGGTACAGCCT  309417

Exon 2
Identities = 68/85 (80%), Gaps = 2/85 (2%) Strand=Plus/Minus

Query  1059    TATTATTTGCACACTGATGAACATGGCATC--TGCTGGCCTTCAGGCAATCCTGTAATGT  1116
               |||||| | || ||| |||||||||| |||  ||| ||||||||| || |||| |||||
Sbjct  306850  TATTATCTCCAAACTCATGAACATGGAATCCATGCAGGCCTTCAGTCAGTCCTTTAATGC  306791

Query  1117    GTCCCAACACCACAACTGGTCCATT  1141
                |||| |||||   ||||||||||
Sbjct  306790  TTCCCCACACCTTCACTGGTCCATT  306766
```

Fig. 18: Alignment of human chromosome 8 (70000-72000) gi:21307426 with Human chromosome 21 gi 7768676

```
70621 ctacataaga agtttctgg**a gaattgt**gaa tgtgtg**tgtg ttttttttt**t ctttacactt
70681 ttaagattac agtgttaaat gtttatcaag ttacatttct acagctagca ggataacaca
70741 tgtgttactg gcactcagac ctcatcctag ggaatgttta ctgggaattc ccaaatacac
70801 cacaagccaa ataacttcca gcatttccca tctttgtact ttcttcattt ccatctttgc
70861 ccaagtcatc tggtttttca tggaccacca ttgcatggcc aatgatggag tggcctcctg
70921 agagtgagat catagaatct tcaatagaca catcattgca tcatctttgc cagcagtcac
70981 attgcccagg tgtttaacat gcctctcttg atcctttggt ccaccgtgtt ttggggatag
71041 aggattaaag tgaggacctg cactggtaca gccttgtata ttatttgcac actgatgaac
71101 atggcatctg ctggccttca ggcaatcctg taatgtgtcc caacaccaca actggtccat
71161 tttccttctg cttgaaatgg ataattctcg gtaccagccc ttcgcccttc agcatgcact
71221 gatggggcca cgctcctgtc tgggttctga ggattgccag ggaaacagga ggtgctataa
71281 gagatgacaa cacaaacgcc **agaattgt**ta atgttttcac tgtccatgat gtatttc**tat**
71341 **ttat**cagagt cctggggtcg ctttgcttgt tgttagataa aagcagtata gtgttatcag
```

Fig. 19: Complete sequences of Cu/Zn processed pseudogene on human chromosome 8 gi:21307462, The 5 and 3' end (agaattgt), a 3' poly (T) rich stretch and is flanked by short direct repeat sequences and TATTTAT(-29) upstream from the 5' direct repeat are shown

(from nt sequence 70671 to 71338) has 98% alignment and a small fragment of 31 nt (from nt coordinate 70626 to 70656) has 100% identity (Fig. 17). The 14 bases (from 70657 to 70670) which do not align with this SOD1 processed pseudogene was found to be a poly T sequence (Fig. 19).

Direct comparison of the complete DNA nucleotide sequences of the SOD1 Ψ 71.4 pseudogene on human chromosome 8 with the SOD1 functional gene on human chromosome 21 was carried out. The whole sequence of this gene has high identities with exon 2, 3, 4 (80, 94 and 84%, respectively) and partially with exon 5 (88%) and no alignment with exon 1 of SOD1 gene on human chromosome 21 (Fig. 18). The complete DNA nucleotide sequence (713 bases) of the Ψ 71.4 SOD1 pseudogene on human chromosome 8 is represented in Fig. 19.

DISCUSSION

NCBI-Blast analysis of the sequence alignment shows that the sequence of SOD1 amplified PCR product of genomic DNA and of SOD1 amplified cDNA are 97% identical. The amplicon of the cattle genomic DNA is 98% identical to mRNA of Bos taurus SOD1 sequences of the Gene Bank database. It also shares sequence homology with mRNA from other species such as *Homo sapiens* (87%), Capra hircus (95%) and *Mucaca mulatto* (91%). This indicates that this gene is an intronless.

The amplicon sequence of the native cattle genomic DNA aligns with four separate segments of Bos taurus chromosome 1 genomic. This findings indicates the presence of SOD1 normal gene i.e., gene with introns and exons on cattle chromosome BTA1. Superoxide dismutase 1 (SOD1) was mapped to cattle chromosome 1q12->q14 by in situ methods (Schmutz *et al*., 1996). It also shares sequence homology (99%) with Bos taurus chromosome BTA13 genomic similar to Superoxide dismutase. This indicates the presence of an intronless (processed) pseudogene on cattle chromosome BTA13. This finding agrees with Gallagher *et al*. (1999) who assigned bovine superoxide dismutase pseudogene to cattle chromosome BTA13. These results indicate that the gene under investigation is an intronless pseudogene. Pseudogenes fall into two distinct categories depending on the mechanism by which they are generated: Processed pseudogenes are reverse transcribed from mRNAs (and thus do not contain introns) whereas nonprocessed pseudogenes arise from duplications of genomic DNA (Mighell *et al*., 2000; Harrison *et al*., 2002a).

Four interesting alignments were found; alignment of the SOD1 amplicon cattle genomic DNA fragment showed 78% identities with human SOD1 processed pseudogene Ψ 69.1; 84% alignment with processed pseudogene Ψ 71.4; 80% alignment with human chromosome 16 DNA and 84% alignment with chromosome 8. The presence of SOD1 pseudogenes has been reported earlier by Danciger *et al*. (1986) who reported four functional processed pseudogenes not residing on chromosome 21. Three of them (69.1, 71.4 and Ψ J) originated from the 0.7-kilobase SOD1 mRNA, while the fourth (Ψ A1) was derived from the 0.9-kilobase mRNA species.

Study of the alignment of the two processed pseudogenes with chromosomes 8 and 16 reveals that processed pseudogene Ψ 69.1 (gi: 181206) is 94% identical with chromosome 16 and only 80% with chromosome 8. Whereas the processed pseudogene Ψ 71.4 (gi: 180712) is 98% identical with chromosome 8. These results indicate that SOD1 processed pseudogene Ψ 69.1 can be assigned to human chromosome 16 and SOD1 processed pseudogene Ψ 71.4 (gi: 180712) can be assigned to chromosome 8. Human chromosome 16 has been reported earlier to be associated with FALS (Sapp *et al*., 2003; Ruddy *et al*., 2003). As well as Messer *et al*. (1992) mapped the motor neuron degeneration gene in the mouse to proximal chromosome 8. He suggested that examination of human chromosome 8, which shows homology of synteny, in human kindreds with Familial Amyotrophic Lateral Sclerosis (FALS) as well as related hereditary neurologic diseases, might be fruitful. These results contradict a study carried out by Huret *et al*. (1987), using in situ hybridization which confirmed that SOD-1 gene localized in the segment enclosing the distal part of human chromosome 21 and no significant labeling on other chromosomes.

In order to identify the complete sequence of the SOD1 gene on chromosome 16, a 2000 bp (from nt coordinate 94000 to 96000) that comprises the human segment of chromosome 16 (from nt coordinate 94952 to 95227) is tested for alignment with the SOD1 processed pseudogenes (gi: 181206). Alignment with the processed pseudogene Ψ 69.1 (gi: 181206) shows that a segment 713 bp (from nt coordinate 94713 to 95423) has 94% alignment and a small fragment of 36 bp (from nt coordinate 94647 to 94682) has 100% alignment. The 30 bp (from nt coordinate 94683 to 94712) which does not align with this processed SOD1 pseudogene found to be poly A (Fig. 16). The results revealed that SOD1 retropseudogene is 777 bases. The sequence homology between the 2000 bp segment of human chromosome 16 and the SOD1 processed pseudogene Ψ 69.1 indicates that the SOD1 processed pseudogene Ψ 69.1 can be assigned to human chromosome 16. This agrees with the results of Sapp *et al*. (2003) that identified a putative locus on chromosome 16q12.

The whole sequence of retropseudogene on human chromosome 16 has identities with exon 2, 3, 4 (82, 86 and 82%, respectively) and partially with exon 5 (88%) and no alignment with exon 1 of SOD1 normal gene on human chromosome 21. This indicates that retropseudogene has many mutations such as deletions and base substitutions in its exons 2, 3, 4 and 5 whereas exon 1 is completely absent. Ruddy *et al.* (2003) suggested mutations of a gene on chromosome 16q12-13 are the significant cause of FALS. As well as Andersen *et al.* (2003) and Hough *et al.* (2004) stated that more than 90 point mutations in the SOD1 gene had been found to lead to the development of FALS. Kunst (2004) reported that the majority of mutations in SOD1 are missense mutations, with a small percentage of deletion and insertion mutations that result in prematurely terminated SOD1 polypeptides. The expression of a mutant SOD1 polypeptide, with or without residual SOD1 activity, is necessary to cause the FALS.

The complete sequence (777 nt) of SOD1 retropseudogene which is discovered and assigned on human chromosome 16 from 94647 nt to 95423 nt is represented in Fig. 16. This retropseudogene lacks introns, contains a 3' poly (A)-rich stretch (30 nt). Retropseudogenes are typically characterized by a complete lake of intron, the presence of small flanking repeats and a polyadenine tail near the 3' end (Zhang *et al.*, 2003). Interestingly, it also contains 5' (acgtc), 3' (accgtc) direct repeats and TATA box (tataa) upstream from the 5' direct repeat, in a position almost identical to the functional TATA box, indicating that it is intronless functional pseudogenes, supporting Danciger *et al.* (1986) who identified the processed pseudogenes Ψ 69.1 (gi: 181206) as a functional gene. These results also confirmed by Zhang and Gerstein (2004) who reported that in most cases, pseudogenes cannot produce transcripts as a result of functional promoter scarcity. Very rarely, some pseudogenes have either retained or acquired a functional promoter so they can be transcribed. Betràn *et al.* (2002) also suggested that there are a substantial number of processed genes of novel function that also derive from mRNAs of various intron-containing genes.

It was also possible to deduce the complete sequence of the SOD1 processed pseudogene on chromosome 8 (713 bp) from NCBI-Blast analysis of a 2000 nt segment of chromosome 8 with the complete processed pseudogene Ψ 71.4 (gi: 180712). The alignment shows that a 668 bp segment has 98% alignment and a small fragment 31 nt has 100% Identity. The 14 bp which does not align with this processed SOD1 pseudogene found to be poly T.

The whole sequence of the complete DNA nucleotide sequences of the SOD1 gene on human chromosome 8 has high identities with exon 2, 3, 4 and 5 and no alignment with exon 1 of SOD1 functional gene on human chromosome 21. Although the SOD1 processed pseudogene on human chromosome 8 overall sequence homology to the SOD1 coding region on human chromosome 21 was extensive, they contain multiple genetic lesions, such as deletions and base substitutions, that preclude the translation of the normal SOD1 polypeptide. It also bears the structural hallmarks of processed pseudogene; it lacks introns, the 5' and 3' direct repeats are presented. It also contains TATA box (TATTTAT) upstream from the 5' direct repeat in a position almost identical to the functional TATA box (-29). It also contains a 3 poly (T)-rich stretch and is flanked by short direct repeat sequences. This results indicate that it is intronless functional pseudogenes, supporting Danciger *et al.* (1986) who identified the processed pseudogenes Ψ 71.4 (gi: 180712) as a functional gene and Nott *et al.* (2003) who reported that pseudogenes are often extremely conserved and transcriptionally active.

The presence of SOD1 functional retropseudoge on chromosome 16 and 8 reported in this study explains the reported FALS and other diseases in individuals with a normal chromosome 21.

ACKNOWLEDGMENTS

This work was partially supported by the executive programme on scientific and technological cooperation between the academy of scientific research and technology of Egypt and the Hungarian academy of sciences (Project: Natural immunity of cattle and other rural ruminants in Egypt). The author wishes to acknowledge Prof. Dr. Soheir M. El Nahas and Prof. Dr. Hanaa Abd El Sadek Oraby for their great effort in putting this manuscript in its final form and for their useful guidance and support.

REFERENCES

Altschul S.F., F. Stephen, T.L. Maden, A.A. Shaffer, Z. Jinghui, Z. Zhang, W. Miller and D.J. Lipman, 1997. Gapped BLAST and PSI-BLAST a new generation of protein database search programs. Nucleic Acids Res., 25: 3389-3402.

Andersen, P.M., K.B. Sims, W.W. Xin, R. Kiely, G. O'Neill, J. Ravits, E. Pioro, Y. Harati, R.D. Brower, J.S. Levine, H.U. Heinicke, W. Seltzer, M. Boss and R.H. Jr. Brown, 2003. Sixteen novel mutations in the Cu/Zn superoxide dismutase gene in amyotrophic lateral sclerosis: A decade of discoveries, defects and disputes. Amyotroph Lateral Scler Other Motor Neuron Disord, 4: 62-73.

Betràn, E., W. Wang, L. Jin and M. Long, 2002. Evolution of the phosphoglycerate mutase processed gene in human and chimpanzee revealing the origin of a new primate gene. Mol. Biol. Evol., 19: 654-663.

Blin, N. and D.W. Stafford, 1976. A general method for isolation of high molecular weight DNA from eukaryotes. Nucleic Acid Res., 3: 2303-2308.

Brosius, J., 1999. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. Gene, 238: 115-134.

Chen, C., A.J. Gentles, J. Jurka and S. Karlin, 2002. Genes, pseudogenes and Alu sequence organization across human chromosomes 21 and 22. Proc. Natl. Acad Sci. USA., 99: 2930-2935.

Danciger, E., N. Dafni, Y. Bernstein, Z. Laver-Rudich, A. Neer and Y. Groner, 1986. Human Cu/Zn superoxide dismutase gene family: Molecular structure and characterization of four Cu/Zn superoxide dismutase-related pseudogenes. Proc. Nat. Acad Sci. USA., 83: 3619-3623.

Gallagher, Jr. D.S., J. Schlapfer, J.D. Burzlaff, J.E. Womack, D.M. Stelly, S.K. Davis and J.F. Taylor, 1999. Cytogenetic alignment of the bovine chromosome 13 genome map by fluorescence *in situ* hybridization of human chromosome 10 and 20 comparative markers. Chromosome Res., 7: 115-119.

Gonçalves, I., L. Duret and D. Mouchiroud, 2000. Nature and structure of human genes that generate retropseudogenes. Genome Res., 10: 672-678.

Grubor, B., J.M. Gallup, D.K. Meyerholz, E.C. Crouch, R.B. Evans, K.A. Brogden, H.D. Lehmkuhl and M.R. Ackermann, 2004. Enhancd surfactant protein and defensin mRNA level and reduced viral replication during parainfluenza virus Type 3 pneumonia in Neonatal lamb. Clin. Diagnlab. Immunol., 11: 599-607.

Harrison, P.M., H. Hegyi, S. Balasubramanian, N.M. Luscombe, P. Bertone, N. Echols, T. Johnson and M. Gerstein, 2002a. Molecular fossils in the human genome: Identification and analysis of the pseudogenes in chromosomes 21 and 22. Genome Res., 12: 272-280.

Harrison, P.M., A. Kumar and N. Lang *et al*., 2002b. A question of size: The eukaryotic proteome and the problems in defining it. Nucleic Acids. Res., 30: 1083-1090.

Hough, M.A., J.G. Grossmann, S.V. Antonyuk, R.W. Strange, P.A. Doucette, J.A. Rodriguez, L.J. Whitson, P.J. Hart, L.J. Hayward, J.S. Valentine and S.S. Hasnain, 2004. Dimer dest may result in disease-causing properties: Structures of motor neuron disease mutants. Proc. Nat. Acad. Sci., 101: 5976-5981.

Huret, J.L., J.M. Delabar, F. Marlhens, A. Aurias, A. Nicole, M. Berthier, J. Tanzer and P.M. Sinet, 1987. Down syndrome with duplication of a region of chromosome 21 containing the CuZn superoxide dismutase gene without detectable karyotypic abnormality. Hum. Genet., 75: 251-257.

Kunst, C.B., 2004. Complex genetics of amyotrophic lateral sclerosis. Am. J. Hum. Genet., 756: 933-947.

Lahn, B.T. and D.C. Page, 1999. Retroposition of autosomal mRNA yielded testis-specific gene family on human Y chromosome. Nat. Genet., 21: 429-433.

Lander, E.S., L.M. Linton and B. Birren *et al*., 2001. Initial sequencing and analysis of the human genome. Nature, 409: 860-921.

Lyons, L.A., T.F. Laughlin, N.G. Copeland, N.A. Jenkins, J.E. Womack and S.J. O'Brien, 1997. Comparative anchor tagged sequences CATS for integrative mapping of mammalian genomes. Nature Genet., 15: 47-56.

Messer, A., J. Plummer, P. Maskin, J.M. Coffin and W.N. Frankel, 1992. Mapping of the motor neuron degeneration and gene, a mouse model of amyotrophic lateral sclerosis ALS. Genomics, 13: 797-802.

Mighell, A.J., N.R. Smith, P.A. Robinson and A.F. Markham, 2000. Vertebrate pseudogenes. FEBS Lett., 468: 109-114.

Noor, R., S. Mittal and J. Iqbal, 2002. Superoxide dismutase-applications and relevance to human diseases. Med. Sci. Monit., 8: 210-216.

Nott, A., S.H. Meislin and M.J. Moore, 2003. A quantitative analysis of intron effects on mammalian gene expression. RNA, 9: 607-617.

Ohshima, K., M. Hattori, T. Yada, T. Gojobori, Y. Sakaki and N. Okada, 2003. Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates. Genome Biol., 74: 1-14.

Ruddy, D.M., M.J. Parton, A. Al-Chalabi, C.M. Lewis, C. Vance, B.N. Smith, P.N. Leigh, J.F. Powell, T. Siddique, E.P. Meyjes, F. Baas, V. de Jong and C.E. Shaw, 2003. Two families with familial amyotrophic lateral sclerosis are linked to a novel locus on chromosome 16q. Am. J. Hum. Genet., 73: 390-396.

Sapp, P.C., B.A. Hosler, D. McKenna-Yasek, W. Chin, A. Gann, H. Genise, J. Gorenstein, M. Huang, W. Sailer, M. Scheffler, M. Valesky, J.L. Haines, M. Pericak-Vance, T. Siddique, H.R. Horvitz and R.H. Jr. Brown, 2003. Identification of two novel loci for dominantly inherited familial amyotrophic lateral sclerosis. Am. J. Hum. Genet., 732: 397-403.

Schmutz, S.M., D. Cornwell, J.S. Moker and D.L. Troyer, 1996. Physical mapping of SOD1 to bovine chromosome 1. Cytogenet. Cell Genet., 721: 37-39.

Shih, C. and R.A. Weinberg, 1982. Isolation of transforming sequences from a human bladder carcinoma cell line. Cell, 29: 161-169.

Siddique, T. and H.X. Deng, 1996. Genetics of amyotrophic lateral sclerosis. Hum. Mol. Genet., 5: 1465-1470.

Yeh, R.F., L.P. Lim and C. Burge, 2001. Computational inference of homologous gene structures in the human genome. Genome Res., 11: 803-816.

Zhang, Z., P.M. Harrison, Y. Liu and M. Gerstein, 2003. Millions of years of evolution preserved: A comprehensive catalog of the processed pseudogenes in the human genome. Genome Res., 13: 2541-2558.

Zhang, Z. and M. Gerstein, 2004. Large-scale analysis of pseudogenes in the human genome. Genet. Develop., 14: 328-335.