ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

In Silico Identification and Molecular Characterization of a New Tumor Marker Mucin-Like Protein 1 (*MUCL1*) Gene in *Ovis aries.*

Nevien M Sabry*.

Cell Biology Department, Genetic Engineering and Biotechnology Division, National Research Centre, Giza 12622, Egypt.

ABSTRACT

An important goal of agricultural genomics is to create detailed comparative maps for economically important species of plants and animals. New methods and resources are speeding the development of comparative maps of additional organisms. To predict the position of genes *in silico* the extensive knowledge of the human genome project and other animal projects must be exploiting. Mucin-like 1 (*MUCL1*) gene is known as small breast epithelial mucin (SBEM) which is predicted to code for a low molecular weight glycoprotein with high similarity to membrane-associated mucins, including *MUC1*.Therefore, the goal of the current study was to identify *Ovis aries MUCL1* gene and its molecular characterization based on the database search and comparison with its potential *Homo sapiens* and *Capra hircus* homologue by using comparative mapping tool. *Ovis aries MUCL1* gene revealed that DNA sequence spans approximately 5.5 kb (comprises 4 exons interrupted by three introns), resulting in 493 bp consensus mRNA sequence with 90-amino acid and present on chromosome 3 (OAR3). The phylogenetic tree of DNA, mRNA and amino acid sequences of *MUCL1* gene showed that *Ovis aries MUCL1* was more related with *MUCL1* gene of *Capra hircus* than that of *Homo sapiens* which is in agreement with the known species relationships.

Keywords: *in silico*, comparative mapping, *Ovis aries*, mucin-like 1 gene, tumor marker

Corresponding author

ISSN: 0975-8585

INTRODUCTION

Mucins are large and highly glycosylated multifunctional proteins found on the surface of epithelial tissues lining the respiratory, digestive and urinogenital tracts [1,2]. Mucins are the major component of the mucus that protects underlying epithelial cells from infection, dehydration and physiological or chemical injury [3]. The mucins provide hydrophilic scaffolding, holding other anti-microbial proteins such as lysozyme, transferrin, immunoglobulins, defensins and members of the trefoil factor family at the epithelial surface. Mucins comprise a family of high molecular weight glycoproteins with a large number of O-glycosylated tandem repeat domains varying in number, length and extent of O-glycosylation [4,5]. Protein products of MUC genes have been studied in tumors arising from various organs, including the breast, colon, pancreas and ovary [6,7].

Mucin 1 (*MUC1*) gene is expressed aberrantly in most human breast cancers. Also, *MUC1* is expressed by many hematopoietic cell types, including T-cells [8], β-cells [9], bone marrow mononuclear cells [10], monocytes [9], dendritic cells [9]. However, *MUC1* displays relatively broad expression among epithelial tissues including the colon, breast, pancreas, ovary, prostate, tracheobronchial tree, stomach, and uterus. For this reason, MUC1-derived tumor antigens have relatively poor specificity for individual tumor types, and their clinical utility is limited to monitoring the efficacy of cancer therapy and warning of tumor relapse or malignant spread [11,12]. Mucin-like 1 gene (*MUCL1*) is identified as a putative breast-specific gene and considered to be a promising breast specific marker [13]. Thus the Mucin-like 1 gene (*MUCL1*) gene is also known as small breast epithelial mucin (SBEM) [14].

The small breast epithelial mucin (SBEM) gene is predicted to code for a low molecular weight glycoprotein with highly similarity to membrane-associated mucins, including *MUC1*. SBEM gene, also known as BS106 and/or B511S, was originally identified as a putative breast-specific gene [13,15,16]. SBEM product is similar to proteins B511S [17] and BS106 [15]. SBEM encodes a secreted 90 amino acids glycoprotein and exhibits characteristics of members of the mucin family [13]. SBEM gene has shown more specific patterns of expression, limited to breast and salivary glands [14].

Parallels between SBEM and known epithelial mucins such as *MUC1*, together with its more narrowly restricted pattern of expression, suggest that this novel gene represents an attractive candidate for a breast biomarker with potential for cancer diagnostics, as well as being a possible future target for the development of a breast tumor vaccine. Moreover, the absence of SBEM expression in normal lymph node tissue suggests that this gene could also be used to detect breast micrometastases in axillary lymph nodes [13]. Sequencing and annotation of mucin genes is known to be difficult due to the large size and repetitive structure of these molecules; moreover, among species, differences in the mucin gene family have been reported. Comparison between humans and mice, two mammals with a completed annotated genome (http://www.ensembl.org/index.html), has shown that although the majority of mucins are commonly represented in both species, differences for few mucin genes are evident.

An important goal of agricultural genomics is to create detailed comparative maps for economically important species of plants and animals. A variety of new methods and resources are speeding the development of comparative maps of additional organisms. The comparative mapping by annotation and sequence similarity (COMPASS) strategy is shown here to be effective in reliably predicting the position of genes *in silico*, exploiting the extensive knowledge of the human genome project. An important aspect of the COMPASS strategy is that the cost of mapping *in silico* is far less than the cost of RH mapping or linkage mapping. Thus, the expense and effort of mapping large numbers of genes to identify candidate genes for QTL can be minimized. In addition, COMPASS can be used in a highly selective and directed manner to fill in gaps in the RH and comparative maps [18].

Therefore, the goal of the current study was to identify *Ovis aries MUCL1*/SBEM gene and its molecular characterization based on the database search and comparison with its potential *Homo sapiens* (human) and *Capra hircus* (goat) homologue by using comparative mapping tool.

ISSN: 0975-8585

METHODS

One can do a homology search to identify genes, by looking for similar genes in the public databases. In gene finding, sequence similarity can be used in different ways as: 1) A direct comparison of a genomic sequence with databases of expressed sequence tags (ESTs), using programs such as BLASTN, 2) A comparison of a translated genomic sequence with a translated genomic or cDNA sequence using BLASTX can identify similarities among coding region, and 3) Comparison of genomic sequences with homologous genomic sequences from closely related organisms (e.g., sheep vs cattle or goat), using BLAST and multiple alignment programs such as CLUSTALW, to identify conserved regions, which often correspond to coding exons or important transcriptional or splicing signal. Regions of DNA that encode proteins are first transcribed into mRNA and then translated into protein so it is important to determine the correct open reading frame (ORF) of the gene to identify amino acids will be encoded by a gene and protein sequences this was done by using Open Reading Frames program (www.ncbi.nlm.nih.gov/gorf/gorf.html).

BLAST and CLUSTALW programs are widely used for database searches. BLAST is a rapid sequence comparison tool that uses a heuristic approach to construct alignments by optimizing a measure of local similarity [19,20]. Since, BLAST compares protein and nucleotide sequences much faster than dynamic programming methods such as Smith-Waterman and Needleman-Wunsch [21,22]. The CLUSTAL programs are widely used for carrying out automatic multiple alignments of nucleotide or amino acid sequences. The most familiar version is CLUSTALW, which uses a simple text menu system that is portable to more or less all computer systems. CLUSTALX features a graphical user interface and some powerful graphical utilities for aiding the interpretation of alignments and is the preferred version for interactive usage [23]. Also, one common use of ORF is as one piece of evidence to assist in gene prediction [24]. Long ORFs are often used, along with other evidence, to initially identify candidate protein coding regions in a DNA sequence.

In this study, to identify mucin-like 1 (*MUCL1*), the *Ovis aries* genome assembly version Oar_v3.1, was used, accessible through the NCBI database (National Center for Biotechnology Information. http://www.ncbi.nlm.nih.gov/). Also, number of the puplished DNA and mRNA of mucin-like 1 (*MUCL1*) gene of *Homo sapiens* and *Capra hircus* were extracted from the GenBank and EMBL databases (Table 1). The predicted *Ovis aries MUCL1* gene then were subsequently aligned with their putative human and goat homologue to calculate amino acid sequence similarity.

Table 1: Published DNA and mRNA of mucin-like 1 (*MUCL1*) gene

Organism	mRNA	DNA	Chromosome
Homo sapiens MUCL1	NM_058173.2	NC_000012.12 (54854515..54858393)	12
Capra hircus MUCL1	XM_005679839.1	NC_022297 (24516200..24521181)	5

RESULTS

To search for mucin-like 1 (*MUCL1*) gene sequence homologies in *Ovis aries*, the DNA and mRNA sequences of *Homo sapiens* and *Capra hircus*, which published in GenBank and EMBL databases, were aligned with the *Ovis aries* genome assembly version Oar_v3.1, using NCBI BLAST analysis (http://blast.ncbi.nlm.nih.gov/Blast). The sequence alignments of the published *MUCL1* mRNA of *Homo sapiens* and *Capra hircus* with *Ovis aries* genome, uncharacterized locus (uncharacterized LOC101118004) of 493 bp was obtained with 86 and 98% similarity with *MUCL1*-mRNA of *Homo sapiens* and *Capra hircus*, respectively. This 493 bp fragment was submitted to the Genbank database and has been assigned as *Ovis aries* mucin-like 1 (*MUCL1*) mRNA with accession number (LC027272). Also, when the published genomic DNA of *MUCL1* gene was aligned with *Ovis aries* genomes, it was found that the nucleotide segment had the same number of the uncharacterized locus (uncharacterized LOC101118004) obtained by mRNA. The results of these alignments also allowed us to assign *MUCL1* gene sequence homology in *Ovis aries* on chromosome 3 (OAR3) with accession numbers (NC_019460.1) region from 131586121 to 131591624, which gave 71% and 98% similarity with *MUCL1*-DNA of both *Homo sapiens* and *Capra hircus*, respectively. The deduced structure of *MUCL1*-DNA gene was composed of four exons interrupted by three introns varying in size, which contained

ISSN: 0975-8585

273 bp ORF (Fig. 1). Also, the sequence was submitted to Genbank and has been assigned with accession number (LC027273).

CLUSTALW program (http://www.ebi.ac.uk/Tools/clustalw) was used to evaluate the identity and differences in nucleotide sequences between mRNA and DNA sequences of *MUCL1* gene published in database of *Homo sapiens* and *Capra hircus* with the obtained fragment of uncharacterized locus of *Ovis aries*.

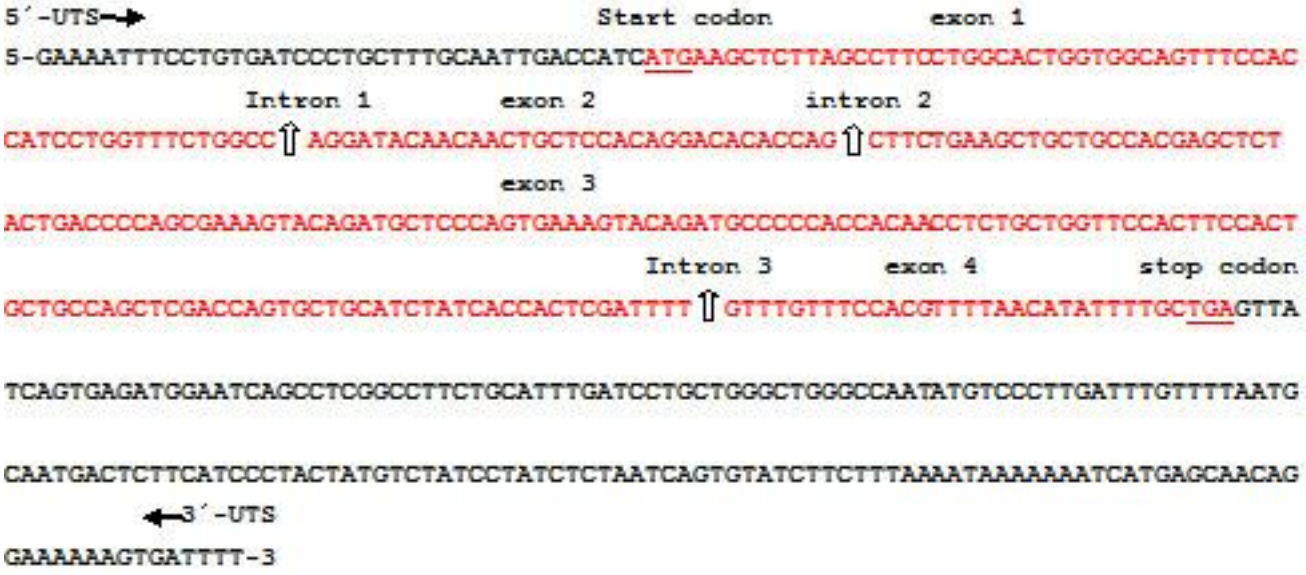


Figure 1: Structure of the sheep mucin-like 1 gene (Sheep *MUCL1*) secreted by *Ovis aries* and has accession number LC027273. Intron 1 was 471 bp, intron 2 was 3367 bp and intron 3 1173 bp.

Multiple nucleotide sequence alignment between mRNA of *Ovis aries* uncharacterized locus with mRNA of *MUCL1* gene for *Homo sapiens* and *Capra hircus* showed 70% and 97% identity, respectively (Table 2 & Fig. 2). Also, multiple nucleotide sequence alignment between the uncharacterized locus DNA of *Ovis aries* with DNA of *MUCL1* gene for *Homo sapiens* and *Capra hircus* showed 71% and 97% identity, respectively (Table 3).

Table 2: Scores of identity for *MUCL1*-mRNA sequences using Clustal2.1

Organisms	*Ovis aries*	*Capra hircus*	*Homo sapiens*
Ovis aries	100		
Capra hircus	97	100	
Homo sapiens	70	70	100

Table 3: Scores of identity for *MUCL1*-DNA sequences using Clustal2.1

Organisms	*Ovis aries*	*Capra hircus*	*Homo sapiens*
Ovis aries	100		
Capra hircus	97	100	
Homo sapiens	71	70	100

```
Ovis aries MUCL1    ------------------------CGAAA-ATTTC---CTGTGATCCC 20
Capra hircus MUCL1  ACTGGAGTCAAGATTCCTCCGGGCTCCGAAA-ATTTC---CTGTGATACC 46
Homo sapiens MUCL1  -CAGCGCCTTGCCTTCTCTTAGGCTTTGAAGCATTTTTGTCTGTGCTCCC 49
                                          ***  ****    ***** * **

Ovis aries MUCL1    TGCTTTGCAATTGACCATCATGAAGCTCTTAGCCTTCCTGGCACTGGTGG 70
Capra hircus MUCL1  TGCTTTGCAATTGACCATCATGAAGCTCTTAGCCTTCCTGACACTGGTGG 96
Homo sapiens MUCL1  TGATCTTCAGGTCACCACCATGAAGTTCTTAGCAGTCCTGGTACTCTTGG 99
                    ** * * **  * **** ******* *******  *****  ***  ***

Ovis aries MUCL1    CAGTTTCCACCATCCTGGTTTCTGGCCAGGAT---ACAACAACTGCTCCA 117
Capra hircus MUCL1  CAGTTTCCACCATCCTGGTTTCTGGCCAGGAT---ACAACAACTGCTCCA 143
```

ISSN: 0975-8585

```
Homo sapiens MUCL1   GAGTTTCCATCTTTCTGGTCTCTGCCCAGAATCCGACAACAGCTGCTCCA 149
                      ******** * * ***** **** **** **   ****** ********

Ovis aries MUCL1     CAGGACACACCAGCTTCTGAAGCTGCTGCCACGAGCTCTACTGACCCCAG 167
Capra hircus MUCL1   CAGGACACACCAGCTTCTGAAGCTGCTGCCACGAGCTCTACTGACCCCAG 193
Homo sapiens MUCL1   GCTGACAC----GTATCC--AGCTACTG------GTCCTGCTGAT---GA 184
                        *****    *  **   **** ***      *  ** ****

Ovis aries MUCL1     CGAAAGTACAGATGCTCCCAGTGAAAGTACAGATGCCCCCACCACAACCT 217
Capra hircus MUCL1   TGAAAGTACAGATGCTCCCAGTGAAAGTACAGATGCCCCCACCACAACCT 243
Homo sapiens MUCL1   TGAAGCCCCTGATGCT------GAAACCACTGCTG------CTGCAACCA 222
                      ***    * ******      ****  ** * **      *  *****

Ovis aries MUCL1     CTGCTGGTTCCACTTCCACTGCTGCCAGCTCGACCAGTGCTGCATCTATC 267
Capra hircus MUCL1   CTGCTGGTTCCACTTCCACTGCTGCCAGCTCGACCAGTGCTGCATCTACC 293
Homo sapiens MUCL1   CTGCAA---CCACTGCTGCTCCTACCACTGCAACCACCGCTGCTTCTACC 269
                      ****     ***** *  ** ** ***   * ****  ***** **** *

Ovis aries MUCL1     ACCACTCG-------ATTTTGTTTGTTTCCACGT---TTTAACATATTTT 307
Capra hircus MUCL1   ACCACTCG-------ATTTTGTTTGTTTCCACGT---TTTAACATATTTT 333
Homo sapiens MUCL1   ACTGCTCGTAAAGACATTCCAGTTTTACCCAAATGGGTTGGGGATCTCCC 319
                      **  ****       ***    ** *  ***  *   **    ** *

Ovis aries MUCL1     G-----CTGAGT-TATCAGTGAGATGGAATCAGCCTCGGCCTTCTGCATT 351
Capra hircus MUCL1   G-----CTGAGT-TATCAGTGAGATGGAATCAGCCTCGGCCTTCTGCATT 377
Homo sapiens MUCL1   GAATGGTAGAGTGTGTCCCTGAGATGGAATCAGCTTGAGTCTTCTGCAAT 369
                      *       **** * **  *************** *  * ******** *

Ovis aries MUCL1     TGATCCTGCTGGGCTGGGCCAATATGTCCCTTGATTTGTTTTAATGCAAT 401
Capra hircus MUCL1   TGATCCTGCTGGGCTGGGCCAATATGTCCCTTGATTTGTTTTAATGCAAT 427
Homo sapiens MUCL1   TGGTCACA----ACTATTC----ATGCTTCCTG---TGATTTCATCCAAC 408
                      ** **        **   *    ***   * **   ** *** ** ***

Ovis aries MUCL1     GACTCTTCATCCCTACTATGTCTATCCTATCTCTAATCAGTGTATCTTCT 451
Capra hircus MUCL1   GACTTTCCATCCCTACTATGTCTATCCTATCTCTAATCAGTGTACCTTCT 477
Homo sapiens MUCL1   TACTTACCTTGCCTACGATATCCCCTTTATCTCTAATCAGTTTATTTTCT 458
                      ***   * * ***** ** **     ************** **  ****

Ovis aries MUCL1     TTAAAATAAAAAA--ATCATGAGCAACAGGAAAAAAGTGATTTT 493
Capra hircus MUCL1   TTAAAATAAAAAA--ATCAGGAGCAACATGAAAAAAGTGA---- 515
Homo sapiens MUCL1   TTCAAATAAAAAATAACTATGAGCAACATAAAAAAAAAAAAA-- 500
                      ** **********  *  * ********  ******   *
```

Figure 2: Multiple nucleotide sequence alignment of *Ovis aries* uncharacterized locus (LOC101118004) mRNA and published sequences of *Capra hircus* MUCL1 and *Homo sapienes* MUCL1. CDS was coloured by red. CLUSTALW (2.1) program was used.

These sequences alignment results revealed that there are high degree of homology between mRNA *Ovis aries* uncharacterized locus with mRNA of *MUCL1* gene for *Homo sapiens* and *Capra hircus*. Also, DNA of *Ovis aries* uncharacterized locus showed homology with DNA of *MUCL1* gene for *Homo sapiens* and *Capra hircus*.

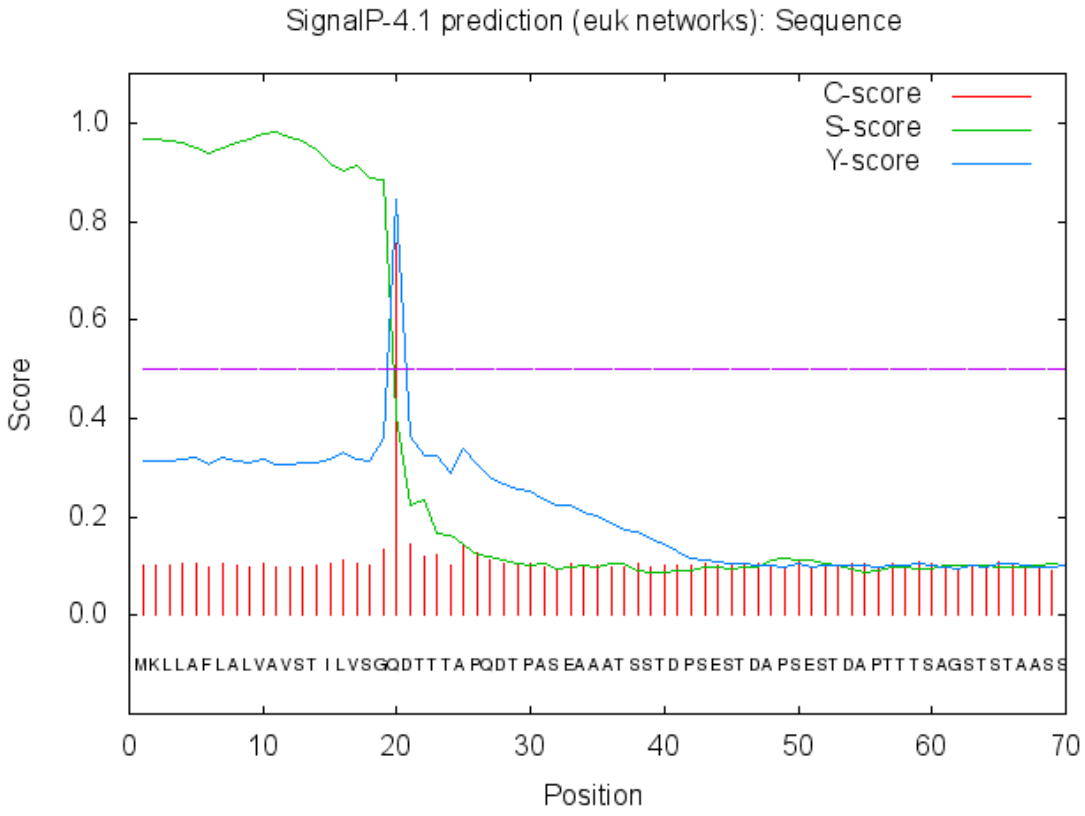
Six Frame Translations of amino acid sequences of each uncharacterized locus *Ovis aries* was carried out using the six frame translation proteins site (http://molbiol.ru/eng/scripts/01_13.html) and/or Open Reading Frame Finder (ORF) site (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). This sequences which are the most relevant frames was aligned with the amino acid sequences of the corresponding MUCL1 gene published using CLUSTALW program. For the amino acid, the amino acid sequence alignment between uncharacterized locus of *Ovis aries* with amino acid of *MUCL1* gene for *Homo sapiens* and *Capra hircus MUCL1* gene showed 23% and 98% similarity, respectively (Table 4 & Fig. 4). However, the amino acid alignment results indicate that the amino acid sequence of *Ovis aries* uncharacterized locus more likely related to amino acid sequence of *Homo sapiens MUCL1* gene also more likely related and share one or more functional domains with amino acid sequence of MUCL1 gene of *Capra hircus*.

ISSN: 0975-8585

Table 4: Scores of similarity for MUCL1 amino acid (aa) sequences using Clustal2.1

	Ovis aries	*Capra hircus*	*Homo sapiens*
Ovis aries	100		
Capra hircus	98	100	
Homo sapiens	23	29	100

The phylogenetic tree of DNA, mRNA and aa sequences of *MUCL1* gene in the different organisms showed that uncharacterized locus of *Ovis aries* was most related with *MUCL1* gene of *Capra hircus* than that of *Homo sapiens* (Figs. 5-7). This clustering based on both of nucleotide and amino acid sequences of *MUCL1* gene clearly showed that the phylogenetic inter-relationship among these species, and also is generally in agreement with the known species relationships.



Measure	Position	Value	Cutoff	Signal peptide?
max. C	20	0.754		
max. Y	20	0.846		
max. S	11	0.980		
mean S	1-19	0.946		
D	1-19	0.900	0.450	YES

Figure 3: The signal peptide length was determined using online SignalP-4.1 euk predictions (http://www.cbs.dtu.dk/services/SignalP/).

ISSN: 0975-8585

```
Ovis aries MUCL1      MKLLAFLALVAVSTILVSGQD-TTTAPQDT-PASEAAATSSTDPSESTDA 48
Capra hircus MUCL1    MKLLAFLTLVAVSTILVSGQD-TTTAPQDT-PASEAAATSSTDPSESTDA 48
Homo sapiens MUCL1    MKFLAVLVLLGVSIFLVSAQNPTTAAPADTYPATGPADDEAPDAETTAAA 50
                      **:**.*.*:.** :***.*: **:** ** **: .*  .:.*.. :: *

Ovis aries MUCL1      PSESTDAPTTTSAGSTSTAASSTSAASITTRFCLFPRFNIFC- 90
Capra hircus MUCL1    PSESTDAPTTTSAGSTSTAASSTSAASTTTRFCLFPRFNIFC- 90
Homo sapiens MUCL1    TTATTAAPTTATTAASTTAR--KDIPVLPKWVGDLPN-GRVCP 90
                      .: :* ****::::.:::**   .. .  .. .  :*. . .*
```

Figure 4: Multiple aa sequences alignment of *Ovis aries* with accession number and published aa sequences of *Capra hircus* MUCL1 and *Homo sapienes* MUCL1. Identical "*", conserved substitution ":", semi-conserved substitutions ".", blank indicates no match " ". CLUSTALW (2.1) program was used.

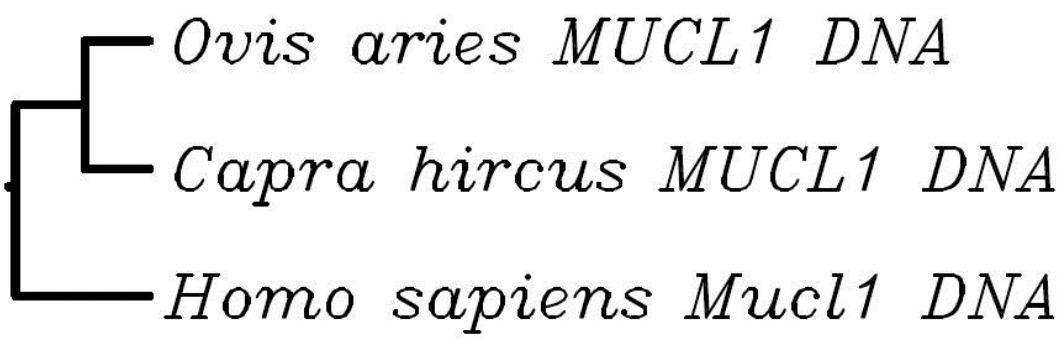


Figure 5: Phylogenetic tree of *MUCL1*-DNA gene showing the relationship between the uncharacterized locus of *Ovis aries* to the published sequences of *Capra hircus* and *Homo sapiens*.

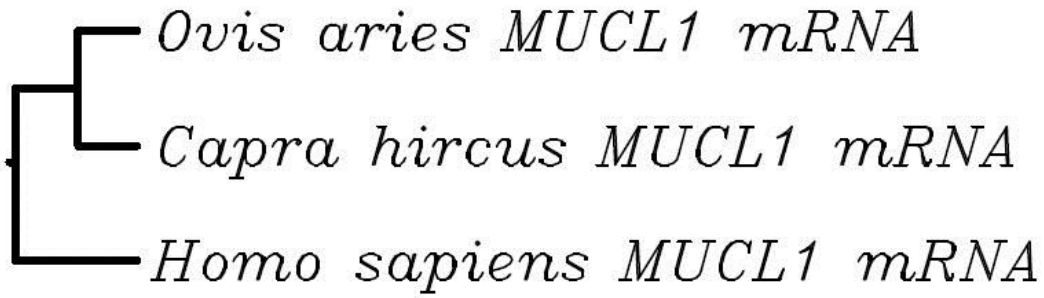


Figure 6: Phylogenetic tree of *MUCL1*-mRNA gene showing the relationship between the uncharacterized locus of *Ovis aries* to the published sequences of *Capra hircus* and *Homo sapiens*.

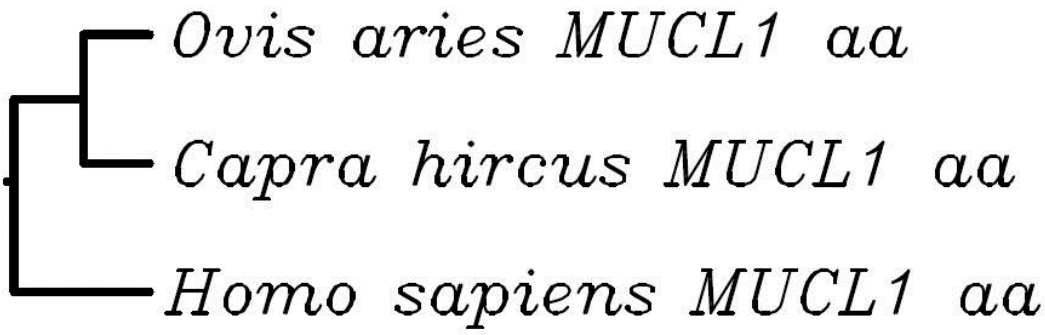


Figure 7: Phylogenetic tree of MUCL1 aa sequence showing the relationship between the uncharacterized locus of *Ovis aries* to the published sequences of *Capra hircus* and *Homo sapiens*.

DISCUSSION

Comparative mapping has been an important tool for gene prediction and the analysis of simple and complex traits in animals [25], and the evolution of the vertebrate genomes [26]. Comparative maps can also identify orthologous regions between animals and human/mouse disease loci [25]. These can be used to define candidate genes for both simple and complex traits. Comparative mapping has been an important tool for the investigation of cattle, sheep and goat genomes [27]. This approach has been combined with the collection of large numbers of gene tags within EST programmes. COMPASS uses homologous DNA sequences (eg ESTs) to search for sequence similarity and putative orthologues in the genome of another species (eg human), and then predicts the chromosome location on the basis of existing comparative maps. This has been pioneered in the creation of high resolution comparative maps between cattle and human [18,27].

The COMPASS is shown to be effective in reliably predicting the position of genes *in silico*, exploiting the extensive knowledge of the human genome project. An important aspect of the COMPASS strategy is that the cost of mapping *in silico* is far less than the cost of RH mapping or linkage mapping. Thus, the expense and effort of mapping large numbers of genes to identify candidate genes for QTL can be minimized. In addition, COMPASS can be used in a highly selective and directed manner to fill in gaps in the RH and comparative maps [18]. Also, the COMPASS permits the prediction of map location of a randomly chosen DNA sequence, provided there is detailed sequence and mapping data for a reference species and adequate existing comparative mapping information for the target species.

Various biological markers have been proposed for the detection of breast cancer cells, including cytokeratin 19, mucin-1, mammaglobin [28,29,30]. However, the frequency of expression of these markers is often related to tumor differentiation and is not always confined to breast tissues. SBEM is identified as a putative breast-specific gene and considered to be a promising breast specific marker [13]. Although published studies have suggested that SBEM might represent a suitable marker for molecular detection of isolated tumor cells in bone marrow and targeting bone marrow micrometastasis in breast cancer patients [14,31].

This study reports the identification and molucular characterization of a DNA and mRNA encoding a novel mucin-like protein1 (*MUCL1*) in *Ovis aris*. The sequence alignments of published mRNA and DNA *MUCL1* gene of *Homo sapiens* and *Capra hircus* with *Ovis aries* genome; uncharacterized locus (uncharacterized LOC101118004) was obtained. The results of these alignments also allowed us to assign this uncharacterized locus LOC101118004 which represents *Ovis aries MUCL1* gene sequence homologies on chromosome 3 (OAR3). Since *MUCL1* gene was mapped to *Homo sapiens* chromosome 12, *Bos taurus* chromosome 5 (BTA5) and *Capra hircus* chromosome 5 (www.ncbi.nlm.nih.gov/gene). Based on the genetic conservation between human chromosome 12 HSA12, cattle chromosome 5 (BTA5) [32,33], goat chromosome 5 (CHX5) [34] and sheep chromosome 3 (OAR3) [35] which confirm that this uncharacterized LOC101118004 locus represent *Ovis aries* Mucin-like protein 1 (*MUCL1*) gene. Recent works on assessment of genomic similarity often use molecular genetic markers [36,37]. However, the most accurate measure of similarity (identity) of genomes is only the fact that chromosomes or chromosomal regions contain similar nucleotide sequences arranged in the similar order, regardless of their alleles. Thus, the compared bovid species showed high similarity of genetic composition [34].

Sheep, goat and cattle which have chromosome- band homology, in these closely related species belonging to Bovidae family, which has a high degree of chromosome conservation between its members [38]. However, humans (*Homo sapiens*) belong to Hominidae which is another family of mammals [39] also there are presence of conserved chromosomal segments between humans and Sheep, goats and cattle [32,40].

Multiple nucleotide sequence alignment between mRNA, DNA and amino acid sequences of *Ovis aries* uncharacterized locus LOC101118004 with *MUCL1* gene of *Homo sapiens* and *Capra hircus* revealed that there are high degree of homology between *Ovis aries* uncharacterized locus with *MUCL1* gene for *Capra hircus*, but showed less homology with *MUCL1* gene of *Homo sapiens*.

Since, mammals are the group of vertebrates (animals with backbones) to which human beings belong. Also, Bovidae are biological family of ruminant mammals. Sheep, goats and cattle are three domestic

species of Bovidae family belong to two subfamily, sheep (*Ovis aries*) and goat (*Capra hircus*) are belong to subfamily Caprine [41], while cattle (*Bos taurus*) belong to subfamily Bovini [42,43]. So, sheep is closely related to goat as both are in the subfamily Caprinae than cattle which belong to subfamily Bovini. The high genetic homology (90%) between sheep and goat has been widely recognized [42].

Alignments of selected sequences which gave more than 70% identity indicated that the aligned sequences are related [44]. CLUSTALW pairwise comparison elucidates identities, similarities and differences in amino acid sequences. Similarity percentages more than 30% indicates that the two sequences are likely related whereas more than 50% indicates that the two sequences are likely to share one or more functional domains. However if the similarity is less than 20% it indicates that the two sequences are not likely to be functionally related [44].

In our study alignment of the Ovis aries uncharacterized LOC101118004 (*MUCL1* gene) revealed that DNA sequence spans approximately 5.5 kb (5504 bp) of 493 bp consensus mRNA sequence with 90-amino acid. Further database analysis showed that this sequence comprises 4 exons interrupted by three introns. This results is in agreement with that Homo sapiens *MUCL1* or SBEM consists of DNA sequence spans approximately 5 kb (5504 bp) include 500-bp mRNA sequence containing a 90-amino acid open reading frame, this sequence also consist of 4 exons interrupted by three introns [13]. *Ovis aries* uncharacterized LOC101118004 amino acid sequence showing presence of a hydrophobic signal peptide (residues 1–19) within the protein sequence suggests that *Ovis aries* MUCL1 is a secreted protein subject to proteolytic processing. The NetOGlyc glycosylation algorithm further predicts this protein to be *O*-glycosylated on most of its 16 threonine residues. On the other way Homo sapiens SBEM encodes a secreted 90 amino acids glycoprotein, which consists of a secretion signal peptide (residues 1–19), three tandemly repeated octapeptide motifs (TTAAXTTA) rich in alanine and threonine residues, that represents a probable target for *O*-glycosylation. Consistent with such posttranslational modification is the presence of a well-defined signal peptide, leading to predict that SBEM is likely to be processed at the apical surface of luminal epithelial cells and to be secreted into the alveolar or ductal lumen [13]. SBEM features suggest strong similarity to many mucins, although this protein lacks a transmembrane domain and is substantially shorter than most other known epithelial mucins [45].

In computational biology gene prediction or gene finding refers to the process of identifying the regions of genomic DNA that encode genes. This includes protein-coding genes as well as RNA genes, but may also include prediction of other functional elements such as regulatory regions. Gene finding is one of the first and most important steps in understanding the genome of a species once it has been sequenced [46]. Gene prediction is one of the key steps in Genome annotation, following Sequence assembly, the filtering of non-coding regions and repeat masking [47].

The phylogenetic research is supported by achievements of modern biology, such as the knowledge of structure and functioning of chromosome apparatus. Cytogenetic techniques are an important tool in systematics and phylogeny of mammals [48]. Analysis of the fine structure of chromosomes is widely used in evolutionary genetics. One of approaches in the study of homology of chromosomes and chromosomal regions in various mammalian species is the comparison of differences in their striation [49,50]. In this study the phylogenetic tree of DNA, mRNA and amino acid sequences of *MUCL1* gene in the different organisms showed that *Ovis aries* uncharacterized locus (LOC101118004) was more related with *MUCL1* gene of *Capra hircus* than that of *Homo sapiens*. This clustering based on both of nucleotide and amino acid sequences of MUCL1 clearly showed the phylogenetic inter-relationship among these species, and also are generally in agreement with the known species relationships.

CONCLUSION

Alignment of the *Ovis aries* uncharacterized LOC101118004 led to the construction of 493 bp consensus mRNA sequence containing CDS 273 bp (from 39-312) with 90-amino acid ORF in which the initiating methionine is framed by a nearly perfect consensus motif for translation initiation (5-CCATCATGA-3). The presence of a hydrophobic signal peptide (residues 1–19) within the protein sequence suggests that *Ovis aries MUCL1* is a secreted protein subject to proteolytic processing. The NetOGlyc glycosylation algorithm further predicts this protein to be *O*-glycosylated on most of its 16 threonine residues. Further database analysis showed that this *Ovis aries* uncharacterized LOC101118004 (now *Ovis aries MUCL1*) DNA sequence

spans approximately 5.5 kb (5504 bp) and comprises 4 exons interrupted by three introns, is present on chromosome 3 (OAR3).

REFERENCES

[1] Hollingsworth MA, Swanson BJ. Nat Rev Cancer 2004; 4: 45-60.
[2] Andrianifahanana M, Moniaux N, Batra SK. Biochim Biophys Acta 2006; 1765: 189-222.
[3] Perez-Vilar J, Hill RL. J Biol Chem 1999; 274: 31754-31751.
[4] Fontenot JD, Tjandra N, Bu D, Ho C, Montelaro RC, Finn OJ. Cancer Res 1993; 53: 5386-5394.
[5] Gendler SJ and Spicer AP. Annu Rev Physiol 1995; 57: 607–34.
[6] Hanski C, Hofmeier M, Schmitt-Graff A, Riede E, Hanski ML, Borchard F, Sieber E, Niedobitek F, Foss HD, Stein H, Riecken EO. J Pathol 1997; 182: 385-91.
[7] Dong Y, Walsh MD, Cumming MC, Wright RG, Khoo SK, Parsons PG, McGuckin MA. J Pathol 1997; 183: 311-7.
[8] Agrawal B, Krantz M J, Parker J, Longenecker B M. Cancer Res 1998; 58: 4079-4081.
[9] Wykes M, MacDonald KPA, Tran M, Quin RJ, Xing PX, Gendler SJ, Hart DNJ, McGuckin MA. J Leukocyte Biol 2002; 72:692-701.
[10] Brugger W, Buhring HJ, Grunebach F, Vogel W, Kaul S, Muller R, Brummendorf TH, Ziegler BL, Rappold I, Brossart P, Scheding S, Kanz L. J Clin Oncol 1999; 17:1535-1544
[11] Cheung P, Allis CD, Sassone-Corsi P. Cell 2000; 103: 263-271.
[12] Burchell JM, Mungul A, Papadimitriou J T. J Mamm Gland Biol Neoplasia 2001; 6: 355-364.
[13] Miksicek RJ, Myal Y, Watson PH, Walker C, Murphy LC, Leygue E. Cancer Res. 2002;62: 2736-2740.
[14] Ayerbes MV, Diaz-Prado S, Ayude D, Campelo RG, Iglesias P, Haz M, Medina V, Gallegos I, Quindos M, Aparicio LA. Adv Exp Med Biol 2008; 617: 331-339.
[15] Colpitts T.L, P Billing, E Granados, M Hayden, S Hodges, L Roberts, J. Russell, P. Friedman, S Stroupe. Tumour Biol 2002; 23: 263-278.
[16] Lacroix M. Endo Relat Cancer 2006; 13: 1033-1067.
[17] Houghton RL, Dillon DC, Molesh DA, Zehentner BK, Xu J, Jiang J, Schmidt C, Frudakis A, Repasky E, Maltez FA, Nolasco M, Badaro R, Zhang X, Roche PC, Persing DH, Reed SG. Mol Diagn 2001; 6: 79-91.
[18] Ozawa A, Band MR, Larson JH, Donovan J, Green CA, Womack JE, Lewin HA. PNAS 2000; 8:4150-4155.
[19] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. J Mol Biol 1990; 215: 403-410.
[20] Altschul SF, Madden TL, Schaëįer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Nucleic Acids Res 1997; 25: 3389-3402.
[21] Smith TF and Waterman MS. J Mol Biol 1981; 147: 195-197.
[22] Needleman SB and Wunsch CD. J Mol Biol 1970; 48: 443-453.
[23] Thompson J D, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. Nucleic Acids Res 1997; 25: 4876-4882.
[24] Deonier R, Tavaré S, Waterman M. Computational Genome Analysis: an introduction. Springer-Verlag. 2005; p. 25. ISBN 0-387-98785-1.
[25] Anderson RS and Beaven A E. Aquat Living Resour 2001; 14: 343-349.
[26] Burt DW, Bruley C, Dunn IC, Jones CT, Ramage A, Law AS, Morrice DR, Paton IR, Smith J, Windsor D, Sazanov A, Nature1999; 402: 411–413.
[27] Band MR, Larson JH, Rebeiz M, Green CA, Heyen DW, Donovan J, Windish R, Steining C, Mahyuddin P, Womack JE, Lewin HA. Genome Res 2000; 10: 1359-1368.
[28] Xenidis N, Perraki M, Kafousi M, Apostolaki S, Bolonaki I, Stathopoulou A, Kalbakis K, Androulakis N, Kouroussis C, Pallis T, Christophylakis C, Argyraki K, Lianidou ES, Stathopoulos S, Georgoulias V, Mavroudis D. J Clin Oncol 2006; 24: 3756-3762.
[29] Zieglschmid V, Hollmann C, Böcher O. Crit Rev Clin Lab Sci 2005; 42:155–196.
[30] Ntoulia M, Stathopoulou A, Ignatiadis M, Malamos N, Mavroudis D, Georgoulias V, Lianidou ES. Clin Biochem 2006: 39:879–887.
[31] Ayerbes MV, Díaz PI, Prado SD, Ayude D, Medina V, Haz M, Reboredo M, Antolín S, Calvo L , Aparicio LMA. J Cancer Res Clin Oncol 2009; 135:1185-1195.
[32] Chowdhary BP, Fronicke L, Gustavsson I, Scherthan H. Mamm Genome 1996; 7: 297-302.
[33] Liu P, Jenkins NA, Copeland NG. Genome Res 2003; 13: 476-484.
[34] Ernst LK, Klenovitskii PM, Bagirov VA, Iolchiev BS, Zinovieva NA, Kalashnikov VV, Fisinin VI, Zhilinskii MA. Agr Biol 2013; 2: 63-70.
[35] Jenkins ZA, Henry HM, Galloway SM, Dodds KG and Montgomery GW. Cell Genet 1997; 78:272-274.

[36] Gorelov PV, Kol'tsov DN, Zinov'eva NA, Gladyr' EA. Agr Biol 2011; 6: 37-40.
[37] Gladyr' EA, Zinov'eva NA, Bagirov VA, Amirshoev FS, Volkova VV, Klenovitskii PM, Karpov AP, Ernst LK. Agr Forestr 2012; 8: 58-62.
[38] Othman OE. Biotechnol 2004; 3:119-125.
[39] Groves CP. Wilson, D. E.; Reeder, D. M, eds. Mammal Species of the World (3rd ed.). Baltimore: Johns Hopkins University Press.2005; pp.181–184. ISBN 0-801-88221-4.
[40] Iannuzzi L, Gallagher DS, Di Meo GP. Cytogenetic Cell Genet 1999; 84: 161-163.
[41] Hirst JA, Howick J, Aronson J, Roberts N, Perera R, Koshiaris C, Heneghan C. *PLoS ONE* 9: e98856.
[42] Iannuzzi L, Skow L, Di Meo GP, Gallagher DS, Womack JE. Chromosome Res 1997; 5: 199-2
[43] Ansari HA, Ellison NW, Reader SM, Badaeva ED, Friebe B, Miller TE, Williams WM. Annals of Botany 1999; 83: 199-206.
[44] Thompson JD, Higgins DG, Gibson TJ. Nucleic Acids Res 1994; 22: 4673-4680.
[45] Moniaux N, Escande F, Porchet N, Aubert JP, Batra SK. Front Biosci 2001; 6: 1192-1206.
[46] Sleator RD and Walsh P. Arch Microbiol 2010; 192: 151-155.
[47] Yandell M and Ence D. Nature Reviews Genetics 2012; 13: 329-342.
[48] Klenovitskii PM, Bagirov VA, Iolchiev BS, Dotsev AV. *Dostizheniya nauki i tekhniki APK*, 2003; 10: 17-19.
[49] Bagirov VA, Klenovitskii PM, Nasibov ShN, Iolchiev BS, Zinov'eva NA, Ernst LK, Gusev IV, Kononov VP. Agr Biol 2009; 6: 27-33.
[50] Nasibov ShN, Bagirov VA, Klenovitskii PM, Iolchiev BS, Zinov'eva NA, Voevodin VA, Amirshoev FS. Agr Forestr 2010; 8: 59-62.