*Available online at* www.pelagiaresearchlibrary.com

Pelagia Research Library

Advances in Applied Science Research, 2011, 2 (5):64-76

ISSN: 0976-8610
CODEN (USA): AASRFC

# Cross class evolutionary pattern of ALCAM proteins and their homologs

**Swati Singh[1] and Anoop Chaturvedi[2]**

[1]*Center of Bioinformatics, University of Allahabad, Allahabad, India*
[2]*Department of Statistics, University of Allahabad, Allahabad, India*

---

## ABSTRACT

*Activated leukocyte cell adhesion molecule ALCAM or CD166 is a type I membrane protein that belongs to the Ig superfamily, and it has five extracellular variable type and constant type Ig domains and a single transmembrane domain. Activated leukocyte cell adhesion molecule is expressed in various tissues, cancers, and cancer-initiating cells. Alterations in expression of ALCAM have been reported in several human tumors, and cell adhesion functions have been proposed to explain its association with cancer. The present study has been done on wide range of ALCAM protein sequences and their homologs from various organisms. In order to decipher the phylogenetic blueprints the proteins from higher organisms to the lower hierarchical order of organisms were selected as the kernel of the investigation. The tree was drawn by Maximum parsimony method while using the bootstrapping as a test of inferred phylogeny. The phylogenetic trees were constructed from multiple aligned sequences showing bootstrap values on nodes and species codes on leaves. The analysis of data led to a single most consistent tree. The study endow with a fine idea about the evolutionary pattern of this protein family and their homologs across an extensive array of animal phyla. Moreover the data provide a basis for future functional studies on this vital protein family.*

**Key words:** ALCAM, CD 166, homologs, phylogeny.

---

## INTRODUCTION

Cell adhesion molecules normalize a variety of endothelial cell functions such as exodus, retort to inflammation, and angiogenesis. Activated leukocyte cell adhesion molecule (ALCAM), a limb of the Ig superfamily, has been detected in the primitive subsets of hematopoietic cells and endothelial cells during embryogenesis [**1**]. Activated leukocyte cell adhesion molecule (ALCAM/CD166) is one of the members of a small subgroup of transmembrane lycoproteins in the immunoglobulin superfamily IgSF. The grown-up proteins of this subgroup are structurally characterized by the presence of five extracellular immunoglobulin Ig domains, comprising two

NH2 -terminal, membrane-distal variable –(V-) type and three membrane-proximal constant –(C2)- type Ig folds, followed by one transmembrane region and a short cytoplasmic tail of variable length (Figure1). The consecutive Ig domains define the $VVC_2C2C_2$ motif characteristic of this subgroup [**2**].


**Figure 1: Schematic representation of the polypeptide structure of the cell adhesion molecules of the VVC2C2C2 subgroup in the immunoglobulin superfamily. All known members are type I transmembrane proteins and consist of five extracellular NH2 –terminal immunoglobulin domains (D1, D2, D3, D4, D5), a transmembrane domain (tm), and a relatively short, C- terminal tail (cyt).**

Human ALCAM was initially isolated as a ligand of CD6, a cell surface receptor belonging to the scavenger receptor cysteine rich protein superfamily [**3**]. CD6 is expressed on the surface of mature T cell and chronic B cell lymphocytic leukemia. Expression of CD6 on T cells is up-regulated following activation, and prior information suggest that CD6 modulates T cell receptor signaling [**4**] Therefore, binding of ALCAM to CD6 suggests that ALCAM may be involved in immune and/or inflammatory responses. In addition, ALCAM has been detected in a variety of cell types where it participates in heterophilic interactions with unknown ligands as well as homophilic interactions, suggesting that ALCAM has varied functions [**2, 5**].

Expression of ALCAM has been associated with differentiation state and progression in many tumors [**6,7**]. In addition, ALCAM is a marker of cancer stem cells and its expression at the tumor cell surface has been correlated with shortened survival in colon-rectal cancers [**8**] and with the vertical growth phase of progression in cutaneous melanoma [**9**].

Despite of being conserved and having major role in expression of this protein in cancerous cell, little is understood about the evolution and genomic diversity of ALCAM proteins among different taxa. Objective of this study was to evolutionarily characterize these proteins in higher eukaryotic organisms as well as to carry out comparative analysis of these organisms based on ALCAM protein sequences. In order to decipher the phylogenetic blueprints the proteins from *Homo sapiens* to the lower hierarchical order of organisms like arthropods were selected as the kernel of the investigation.

## MATERIALS AND METHODS

In order to search ALCAM protein family members PSI BLAST [**10**] was performed by using blastp program in the protein database at NCBI [**11**]. This database has also been used by A. Karunasagar et al **[12]**. *Homo sapiens* protein's gi|68163411|ref|NP_001618.2| CD166 antigen precursor was used as seed sequence and the iterations were performed upto level 5. From the hits 36 sequences (**Table 1)** each from different proteins were selected for further studies. All the sequences were taken in FASTA format. The sequences were examined individually and aligned using ClustalW [**13**] similar to Shashidher et al [**14**] . Bioedit version 7.0.9.0 [**15**] was used for manual editing and analysis of sequences. Entropy **[16]** was calculated as

$$H(l) = -\Sigma f(b,l)\log_{(\text{base } 2)} f(b,l)$$

where *H(l)* = the uncertainty, also called *entropy* at position *l*, *b* represents a residue (out of the allowed choices for the sequence in question), and *f(b,l)* is the frequency at which residue *b* is found at position *l*. The information content of a position *l*, then, is defined as a decrease in uncertainty or entropy at that position. A window of defined size that was 13 is moved along a sequence, the hydropathy scores were summed along the window, and the average (the sum divided by the window size) was taken for each position in the sequence. The mean hydrophobicity value was plotted for the middle residue of the window. Eisenberg *et. al.* method [**17**] was used to plot hydrophobic moment profile with a window size of 13 residues having six residues on either side of the current residue and rotation angle, $\theta$ =100 degrees.

$$\mu H = \{[\Sigma H_n \sin(\delta n)]\text{^}2 + [\Sigma H_n \cos(\delta n)]\}\text{^}(1/2),$$

Where $\mu H$ is the hydrophobic moment, *Hn* is the hydrophobicity score of the residue *H* at position *n*, δ=100 degrees, *n* is position within the segment, and each hydrophobic moment is summed over a segment of the same defined window length.

**Table 1: ALCAM protein and related homolog sequences with their length and NCBI accession code.**

| S.No. | Name of Species | Protein | NCBI Accession Code | Lengthaa |
|---|---|---|---|---|
| 1 | *[Homo sapiens]* | CD166 antigen precursor | gi\|68163411\|ref\|NP_001618.2\| | 583 |
| 2 | *[Homo sapiens]* | MEMD | gi\|3183975\|emb\|CAA71256.1\| | 582 |
| 3 | *[Rattus norvegicus]* | CD166 antigen precursor | gi\|13929058\|ref\|NP_113941.1\| | 583 |
| 4 | *[Bos taurus]* | CD166 antigen precursor | gi\|296491434\|gb\|DAA33487.1\| | 583 |
| 5 | *[Mus musculus]* | CD166 antigen precursor | gi\|31791059\|ref\|NP_033785.1\| | 583 |
| 6 | *[Pongo abelii]* | CD166 antigen precursor | gi\|197099250\|ref\|NP_001126619.1\| | 570 |
| 7 | *[Gallus gallus]* | BEN glycoprotein precursor | gi\|63088\|emb\|CAA45579.1\| | 588 |
| 8 | *[Gallus gallus]* | CD166 antigen precursor | gi\|45383998\|ref\|NP_990510.1 | 588 |
| 9 | *[Homo sapiens]* | kin of IRRE-like protein 1 precursor | gi\|166295173\|ref\|NP_060710.3\| | 757 |
| 10 | *[Rattus norvegicus]* | kin of IRRE like 1 Drosophila | gi\|149048212\|gb\|EDM00788.1\| | 603 |
| 11 | *[Mus musculus]* | Kirrel protein | gi\|112180401\|gb\|AAH23765.3\| | 634 |
| 12 | *[Homo sapiens]* | NEPH1 | gi\|14572521\|gb\|AAK00529.1\| | 605 |
| 13 | *[Xenopus laevis]* | ALCAM | gi\|148232286\|ref\|NP_001085996.1\| | 573 |
| 14 | *[Xenopus laevis]* | MGC84135 protein | gi\|148228231\|ref\|NP_001086195.1\| | 570 |
| 15 | *[XenopusSilurana tropicalis]* | ALCAM | gi\|118404578\|ref\|NP_001072753.1\| | 560 |
| 16 | *[Rattus norvegicus]* | NEPH1 | gi\|30314348\|gb\|AAP12626.1\| | 702 |
| 17 | *[Aedes aegypti]* | nephrin | gi\|157131565\|ref\|XP_001655882.1\| | 505 |
| 18 | *[Rattus norvegicus]* | rCG54063 | gi\|149056328\|gb\|EDM07759.1\| | 565 |
| 19 | *[Danio rerio]* | activated leukocyte cell adhesion molecule | gi\|94732958\|emb\|CAK05472.1\| | 544 |
| 20 | *[Homo sapiens]* | MUC18 glycoprotein | gi\|529724\|gb\|AAA20922.1\| | 646 |
| 21 | *[Homo sapiens]* | Melanoma cell adhesion molecule variant | gi\|62089436\|dbj\|BAD93162.1\| | 659 |
| 22 | *[Mus musculus]* | Bcam protein | gi\|13435987\|gb\|AAH04826.1\| | 650 |
| 23 | *[Mus musculus]* | Lutheran antigen | gi\|10566959\|dbj\|BAB16053.1 | 622 |
| 24 | *[Mus musculus]* | l-gicerin/MUC18 | gi\|10566953\|dbj\|BAB16050.1\| | 648 |
| 25 | *[Homo sapiens]* | B-CAM | gi\|535179\|emb\|CAA56327.1\| | 588 |
| 26 | *[Gallus gallus]* | HEMCAM | gi\|1621230\|emb\|CAA70080.1\| | 626 |
| 27 | *[Mus musculus]* | NEPH1 | gi\|14572519\|gb\|AAK00528.1\| | 392 |
| 28 | *[Mus musculus]* | transmembrane glycoprotein | gi\|452103\|gb\|AAA37528.1\| | 357 |

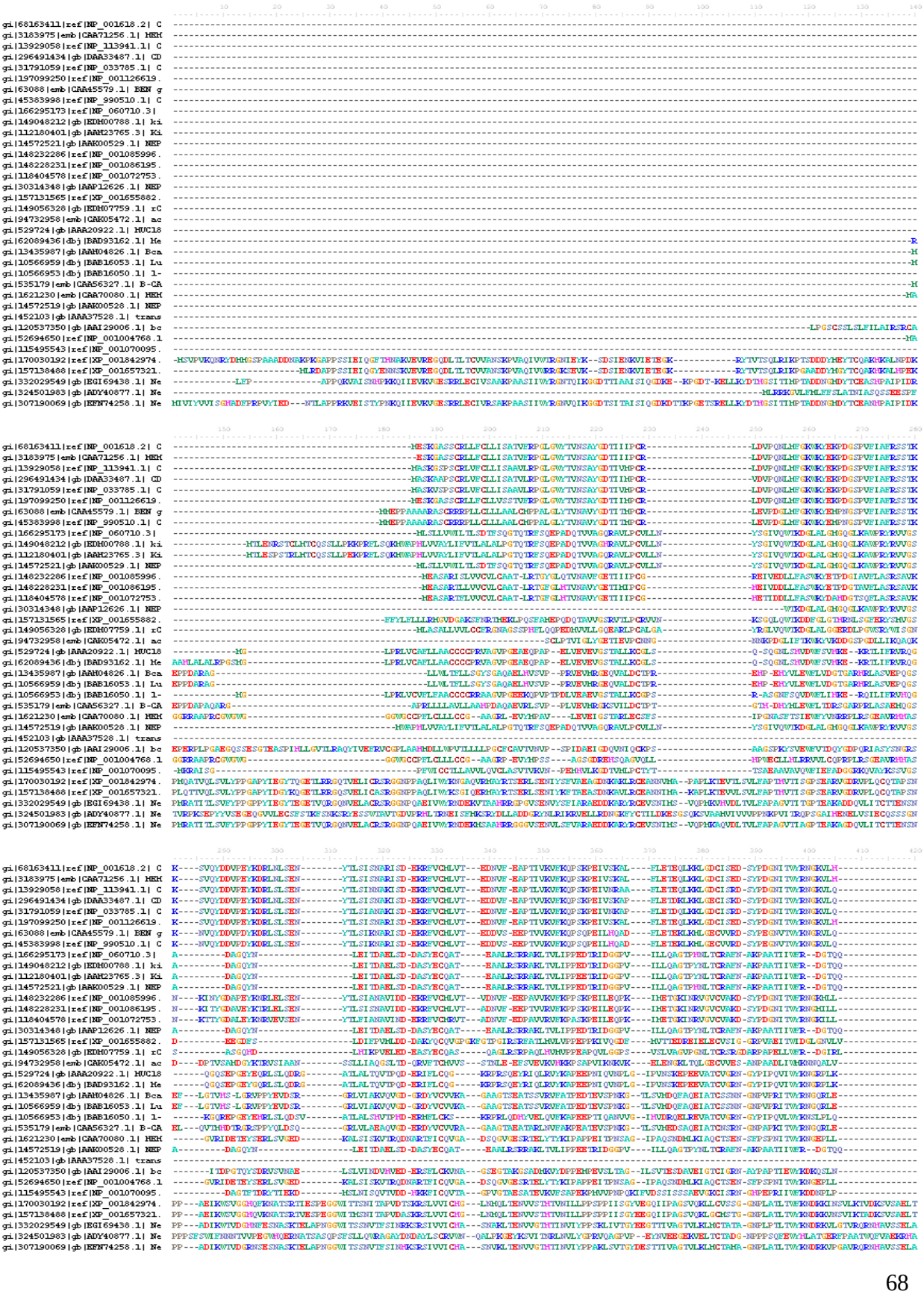
| | | | | |
|---|---|---|---|---|
| 29 | *[Xenopus (Silurana) tropicalis]* | bcam protein | gi\|120537350\|gb\|AAI29006.1\| | 673 |
| 30 | *[Gallus gallus]* | melanoma cell adhesion molecule | gi\|52694650\|ref\|NP_001004768.1\| | 626 |
| 31 | *[Danio rerio]* | basal cell adhesion molecule | gi\|115495543\|ref\|NP_001070095.1\| | 547 |
| 32 | *[Culex quinquefasciatus]* | nephrin | gi\|170030192\|ref\|XP_001842974.1\| | 1360 |
| 33 | *[Aedes aegypti]* | nephrin | gi\|157138488\|ref\|XP_001657321.1\| | 1262 |
| 34 | *[Acromyrmex echinatior]* | Nephrin | gi\|332029549\|gb\|EGI69438.1\| | 1273 |
| 35 | *[Ascaris suum]* | Nephrin | gi\|324501983\|gb\|ADY40877.1\| | 1168 |
| 36 | *[Camponotus floridanus]* | Nephrin | gi\|307190069\|gb\|EFN74258.1\| | 1300 |

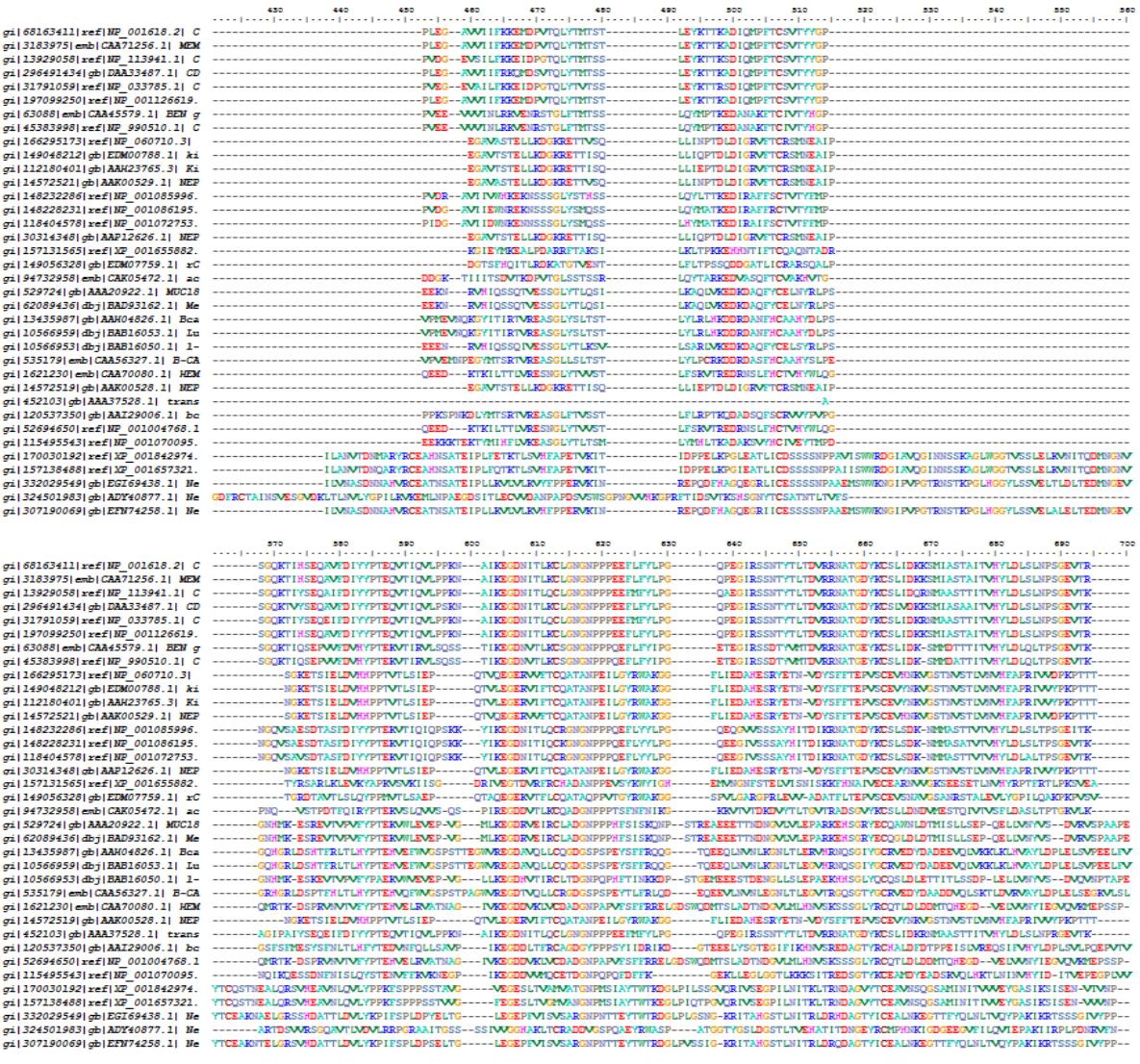
Multiple sequence alignment, phylogenetic and molecular evolutionary analyses were conducted using MEGA version 5 [**18**]. Similar approach has been applied by A. Mani et al**. [19].** For pair wise and multiple alignment gap open penalty was 10 and gap extension penalty was -1 [**20**]. BLOSUM weight matrix was used for substitution scoring [**21**]. The multiple alignments of sequences of ALCAM proteins were used to create phylogenetic trees. The evolutionary history was inferred using the Maximum Parsimony method [**22**]. The consensus tree inferred from 18 most parsimonious trees is shown. Branches corresponding to partitions reproduced in less than 50% trees are collapsed. The consistency index is 0.768402 (0.767846), the retention index is 0.875202 (0.875202), and the composite index is 0.672507 (0.672020) for all sites and parsimony-informative sites. The percentage of parsimonious trees in which the associated taxa clustered together are shown next to the branches. The MP tree was obtained using the Close-Neighbor-Interchange algorithm [**23**] with search level 0 in which the initial trees were obtained with the random addition of sequences (10 replicates). The analysis involved 36 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 220 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 [**18**].

## RESULTS AND DISCUSSION

### 3.1 Multiple sequence alignment

The Multiple alignment of CD166 and related sequences (Figure 2) resulted into an alignment having 1466 positions. By statistical analysis of multiple aligned sequences it was observed that leucine, valine, serine, threonine, glycine, glutamic, alanine, proline, lysine, isoleucine are the most frequently present amino acids with frequency percentage of 8.29, 8.01, 8.47, 7.62, 6.92, 6.67, 6.45, 6.26, 5.46, 5.19 respectively. The multiple aligned sequence of ALCAM and other related homologs protein sequences was found with No. of conserved sites=121, No. of parsimony informative sites= 1028 and No. of singleton sites= 214.





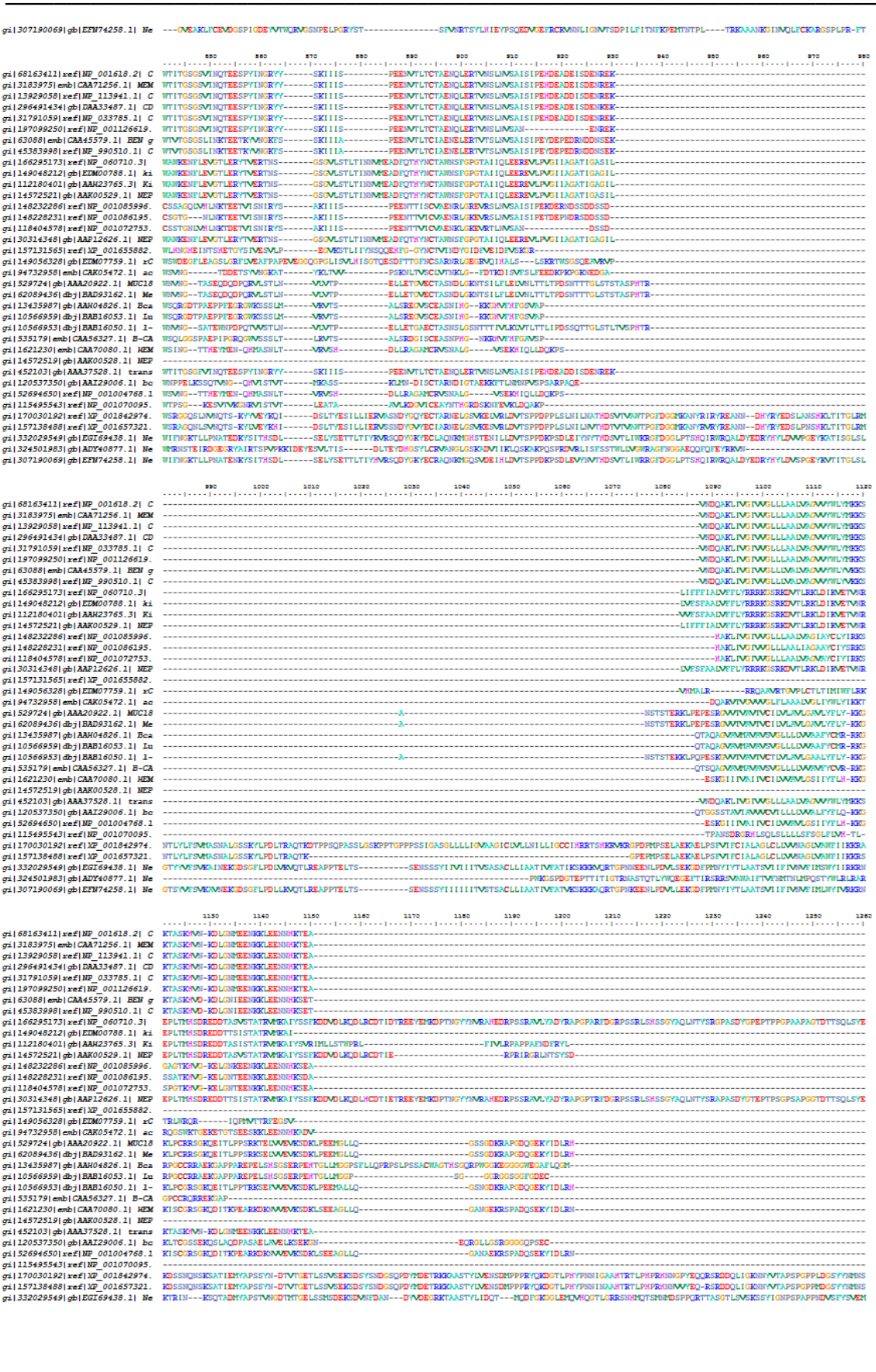


**Figure 2: MSA of ALCAM and related homologs**

## 3.2 Entropy plot

An entropy plot, measure of the lack of the information content and the amount of variability, was generated for all the aligned positions. The plot shows that entropy rarely goes above a scale of two, showing minimal entropy at several positions from position 950 to position 1090 and also from position 1210 to position 1450 where entropy rarely crosses a scale of one, which is a sign of minimal variability in the region.

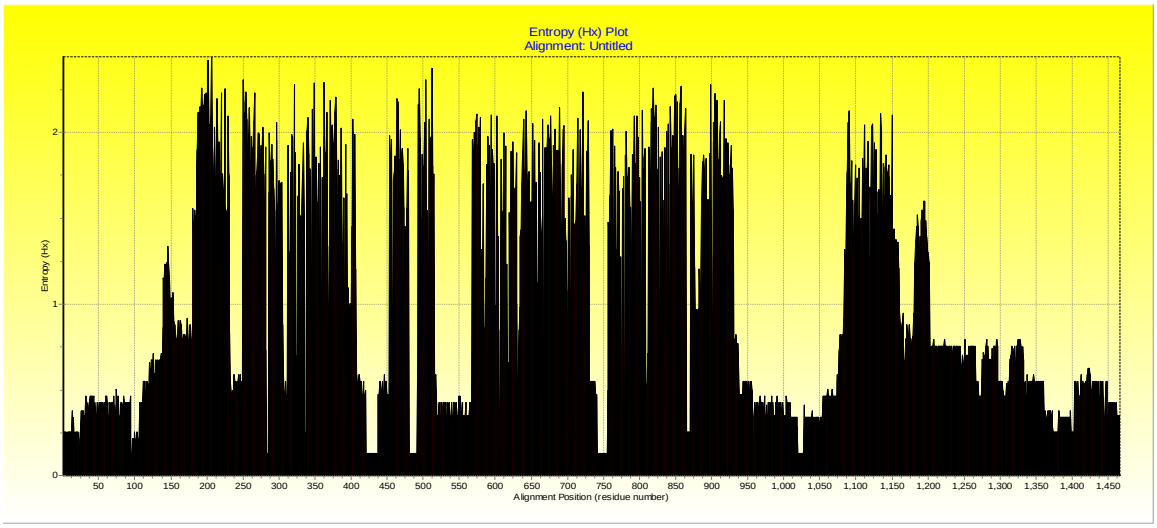



**Figure 3: Entropy (Hx) Plot for aligned ALCAM and related homologs protein sequences**

### 3.3 Hydrophobicity Profile

A hydrophobicity profile plot shows that mean hydrophobicity of the protein for most of the positions is in all the species is below zero, occassionaly it turns to be positive. Maximum hydrophobicity is observed from positions 100 to 200 for nephrin proteins of *Ascaris Suum, Aedes aegypti and Mus musculus.*

From Position 1010 to 1060, increase in hydrophobicity of *Culex quinquefasciatus, Acromyrmex echinatior, Camponotus floridanus* can be seen from Figure 4.a and *HomoSapiens, Musmusculus,* and *Xenopus (Silurana) tropicalis* exhibits high increase in hydrophobicity from position 1060 and 1150. These proteins are basically of non-hydrophobic in nature as most of the positions are across show a below mean hydrophobicity in the case of most of the organisms studied here.

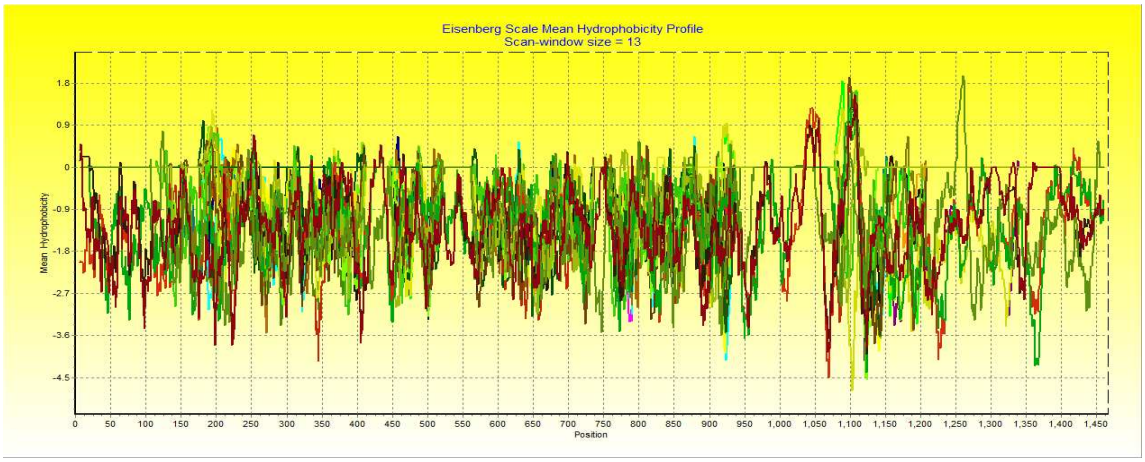



**Figure 4.a: Eisenberg scale mean hydrophobicity profile plot**

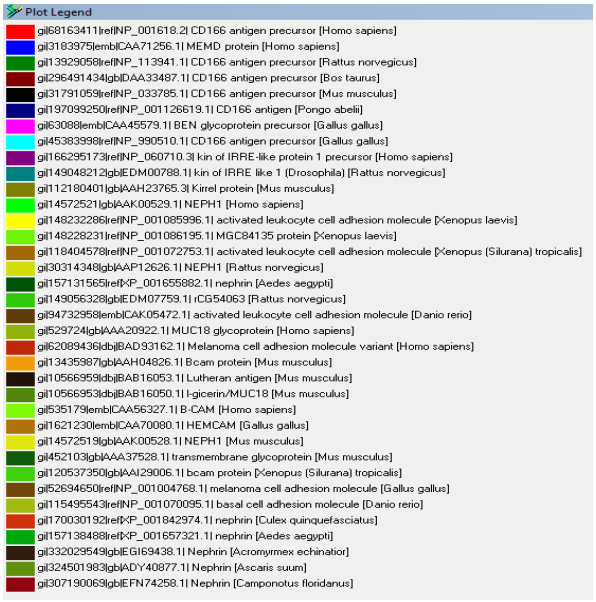



**Figure 4.b: Colour legend for Eisenberg scale mean hydrophobicity profile plot**

### 3.4 Phylogeny

The phylogenetic tree was constructed by using Maximum Parsimony method. The tree shows different organisms on tree nodes branched on the basis of their ALCAM and related homologs proteins. *Nephrin* proteins of *Camponotus floridanus, Acromyrmex echinatior, Aedes aegypti* and *Culex quinquefasciatus* craft a totally diverged branch from the main tree among 36 selected proteins. From a close view at the tree (Figure 5) it becomes clear that the tree can be sub divided mainly into three clusters. Cluster A belongs to hierarchy of nephrin and similar proteins showing evolution of Nematode, arthropod and mammalian nephrins consecutively. However cluster B mainly represents HEMCAM and BCAM proteins and the ALCAM proteins are represented by cluster A. Close homology of Nephrin proteins to the Cell adhesion molecule proteins in sequence similarity search shows their close evolutionary relationship and from the tree it appears that during the course of evolution of cell adhesion molecule proteins BCAM, HEMCAM, MCAM and ALCAM evolved consecutively. ALCAM proteins appear to have evolved rapidly in parallel order to other CAM proteins and specialized themselves as leukocyte cell adhesion molecules during the course of evolution.

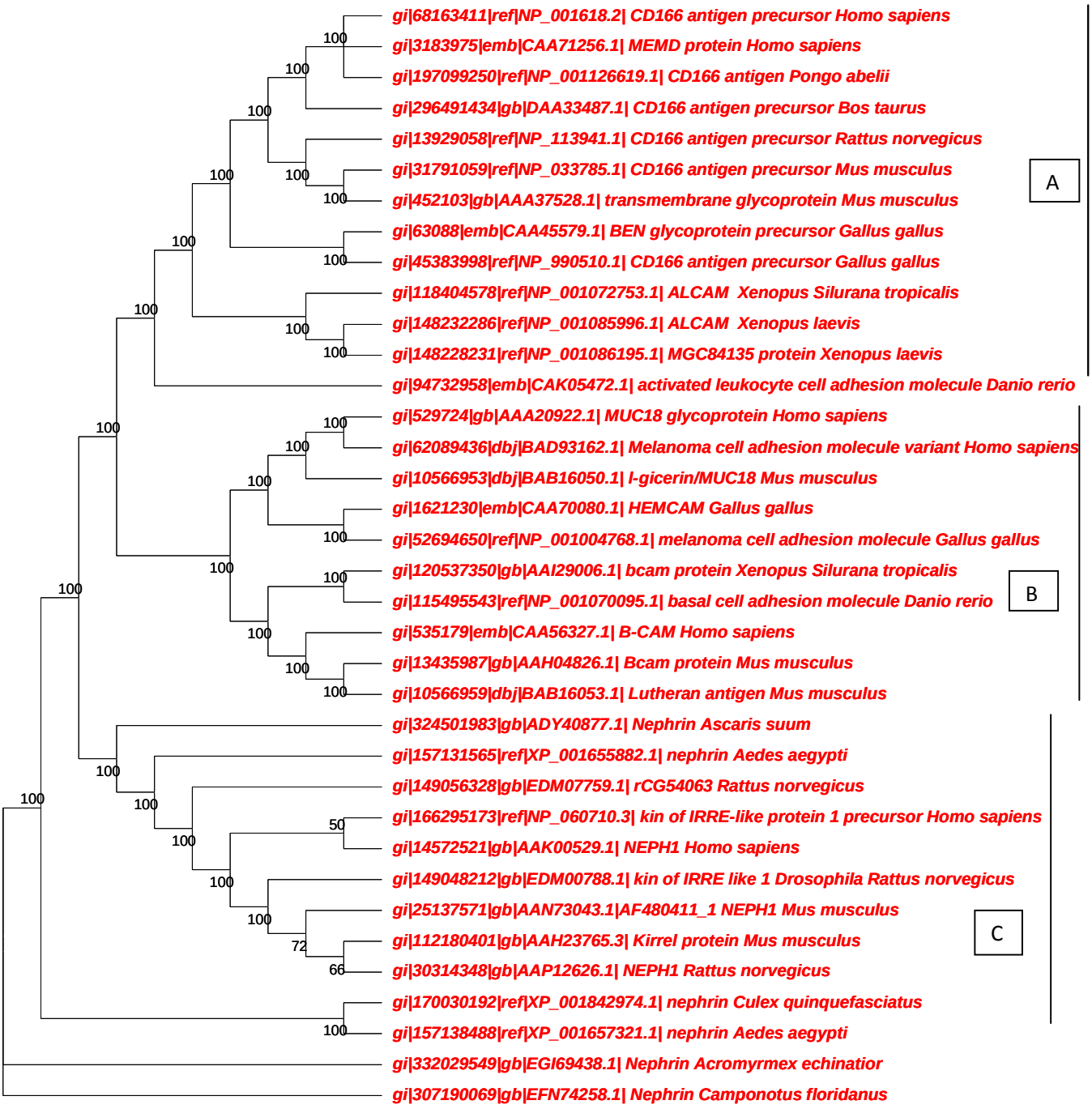



**Figure 5: Phylogenetic tree of ALCAM proteins and their homologs constructed by maximum parsimony method.**

## CONCLUSION

This is the first study that has been done on wide range of ALCAM protein sequences and their homologs from various organisms. In order to decipher the phylogenetic blueprints the protein

sequences from higher organisms to the lower hierarchical order of organisms were selected as the kernel of the investigation. The study endows with a fine idea about the evolutionary pattern of this ALCAM family and other cell adhesion molecule proteins across an extensive array of animal classes. Moreover the data provide a basis for future functional studies on this vital protein family.

**Acknowledgements:**
First Author is thankful to University Grants Commission, New Delhi for a research fellowship.

## REFERENCES

[1] K. Ikeda and T. Quertermous, *J. Biol. Chem. Vol.*, **2004**, 279, No. 53.
[2] G.W. Swart, *European Journal of Cell Biology*, **2002**, 81, 313- 321.
[3] D.D. Patel, S.F. Wee, L.P. Whichard, M.A. Bowen, J.M. Pesando, A. Aruffo, B. F. Haynes, *J Exp Med.*,1995,1814:1563-8.
[4] D.D. Patel, B.F. Haynes, G.C. Starling, J.A. Gebe, and J.Bajorath, *Immunol. Today*, **1997**, 18, 498–504.
[5] A.P. DeBernardo, S. Chang, *J. Cell Biol.*, **1997**, 133, 657–666.
[6] F. Micciche`, L. D. Riva, M. Fabbi, S. Pilotti, P. Mondellini, S. Ferrini, S. Canevari, M.A.Pierotti, I.Bongrazone, *PLoS ONE 62: e17141.*, **1997**, doi:10.1371/journal.pone.0017141.
[7] N.K. Haass, K.S.Smalley, L. Li, M. Herlyn, *Pigment Cell Res*, **2005**, 18: 150-9.
[8] T.G. Levin, A.E. Powell, P.S. Davies, A.D. Silk, A.D.Dismuke *et al*., *Gastroenterology*, **2010**, 139: 2072–82.
[9] L.C. van Kempen, J.J. van den Oord, G.N. van Muijen, U.H. Weidle, H.P. Bloemers, *et al*., *A m J Pathol*, **2000**, 156: 769–74.
[10] S.F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman, *Nucleic Acids Res.*, 2000, 25, 3389-3402.
[11] www.ncbi.nlm.nih.gov/entrez, National Centre for Biotechnology Information.
[12] R. Adiga, I. Karunasagar, I. Karunasagar ,Advances in Applied Science Research, **2010**, 1 (3), 222-228.
[13] D. Higgins, J. Thompson, T. Gibson, J.D. Thompson, D.G. Higgins, T.J. Gibson, *Nucleic Acids Res.*, **1994**, 22, 4673-4680.
[14] B. Shashidher, P. C. Kamal, A. Swetha, A. Priyanka , N. Prudhvi, *Der Pharmacia Sinica*, **2011**, 2 (3): 131-145.
[15] T.A. Hall, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.*, **1999**, 41, 95-98.
[16] J.R. Pierce; an Introduction to Information Theory: Symbols, Signals and Noise, Dover Publications, *Inc., New York*, **1980**.
[17] D.E. Eisenberg, M. Komaromy and R.Wall, *J. Mol. Biol.*, **1984**, 1791, 125-42.
[18] K. Tamura, D. Peterson, N. Peterson, S. Stecher, M. Nei and S. Kumar, *Molecular Biology and Evolution*, *Mol Biol Evol.*, **2011,** May 4.
[19] A. Mani, S. Singh, M. Dwivedi, V. Tripathi, D.K.Gupta, 2011, European Journal of Experimental Biology,2011,1(1), 148-155.
[20] S.F.Altschul, G. Gish, *Methods Enzymol.*, **1996**, 266, 460-480.
[21] S. Henikoff S., J.Henikoff , *Proceedings of the National Academy of Sciences USA* , **1992**,89, 10915-10919.

[22] S. Sridhar**,** F.Lam**,** G.E. Blelloch**,** R. Ravi**,** R. Schwartz, *BMC Bioinformatics,* **2007**, 8**:**472.
[23] M. Nei, S. Kumar; Molecular Evolution and Phylogenetics. *Oxford University Press, New York,* **2000**.