Evolutionary Analytics on Lysosomal Associated Membrane Protein -1 (LAMP-1)

Manish Dwivedi[1], Vijay Tripathi[2], Ashutosh Mani[3] and Dwijendra K. Gupta[4]

Abstract - Lysosomes,the endocytic subcellular compartments play a very important role in the disintegration and recycling of cellular substances. The lysosomal associated membrane proteins LAMP-1 and LAMP-2 are major constituents of the lysosomal membrane. Using different bioinformatics tools, we established the phylogenetic relationship among LAMP-1 proteins from different organisms. The phylogenetic analytics based on ClustalW, MEGA4 and BioEdit softwares showed structural as well as qualitative similarities and dissimilarities of LAMPs and helped us to predict the nature, structure and localization of amino acid of the membrane proteins in the lysosomal membrane. This information can help one explain the molecular basis of different metabolic diseases associated with the lysosomal membrane proteins like Lysosomal storage diseases, I-cell disease etc. and to solve the questions related to biogenesis of lysosomes. This study includes the alignment of the sequences of LAMP-1 by ClustalW and revealed that Ala,, Gly, Leu, Asp and Ser are most frequently occurring amino acids with higher frequency percentage and position from 240 to 420 showed minimal entropy. The entropy rarely touched the scale of two in entropy plot that is a sign of better alignment. It revealed that these proteins were more hydrophobic in N- terminal and C-terminal domain and basically of non-hydrophobic in nature. This work also represented the evolutionary order of LAMP-1 proteins among different mammalians. This study presents the first comparative genomic study and evolutionary analysis of the LAMP-1 proteins across family of organisms with special reference to mammals.

Index Terms -
Endocytic, phylogenetic, LAMP, hydrophobicity, lysosomes.

ABBREVIATIONS

LAMP- Lysosomal associated membrane proteins, NCBI-National Center for Biotechnology Information, BLAST-Basic local alignment search tool, BLOSUM-Blocks of amino acid substitution matrix. MEGA- Molecular Evolutionary Genetics Analysis.

1. INTRODUCTION

Lysosomes represent membrane-bound dense organelles of eukaryotic cells, specialized in breakdown of all four classes of macromolecules. Materials delivered to lysosomes by endocytosis / phagocytosis or autophagocytosis are degraded in these organelles by concerted action of more than 40 hydrolases. The degradation products are then exported to the cytosol through specific transporters and reused in the cellular metabolism [6].

[1,2,3,4]*Center of Bioinformatics, University of Allahabad, Allahabad- 211002, India*
E-Mail: [4]dwijenkumar@rediffmail.com and
[1]mdwivedibio@yahoo.com

The physiological importance of lysosomal metabolite efflux is illustrated by the existence of a group of lysosomal storage diseases with transport defects, such as sialic acid storage disorders and nephropathic cystinosis. These inherited diseases result from defective efflux of sialic acid and cystine from lysosomes, respectively, and they have been linked to mutations in the membrane proteins sialin and cystinosin [20, 21], which are believed to represent sialic acid and cystine transporters. However, most lysosomal transporters, although biochemically characterized, remain unknown at the molecular level.

After the discovery of lysosomes by de Duve and coworkers [3] and conceptualization of inborn lysosomal disease' by Hers, a wide interest in understanding the biology and pathology of lysosomal disorders has led to the discoveries of nearly all lysosomal hydrolases and their encoding genes. However, the knowledge about proteins of the lysosomal membrane, which controls the interchange with other compartments and the cytosol and restricts the aggressive enzymes to the lysosomal interior, remained incomplete [4]. More than 20 distinct transport processes facilitating mainly the export of degradation products across the lysosomal membrane have been characterized functionally [15]. Yet, most of the transport catalyts remain unknown. Similarly, biogenesis of lysosomes and the machinery regulating their interaction with other compartments are still incompletely understood. In a notable report on rat tritosomal membranes, 219 proteins were identified, including 24 novel tentatively lysosomal proteins [14].

Two major lysosomal membrane sialoglycoproteins with apparent Mr ~ 120,000 containing polylactosaminoglycan comprise approximately 0.1-0.2% of total cell proteins. Immunoelectron microscopic examination of HeLa cells localized these two glycoproteins mainly to lysosomes and multivesicular bodies [4]. A number of different cell lines also express these glycoproteins. However, the apparent molecular weights differed between cell lines probably due to differences in the amount of polylactosaminoglycan expressed by each cell line. Fukuda(1988) reported that one of the glycoproteins is very homologous to that of a mouse counterpart, m-lamp- 1 The analogous human form of this glycoprotein is named human lamp-1 (h-lamp-1), while the other glycoprotein, to which the monoclonal antibody was made, is called human lamp-2 (h-lamp-2). Polylactosaminoglycans are heterogenous saccharides often having high molecular weights [18] and represent various antigenic structures such as AB0 blood group antigens, developmental antigens such as mouse F9 antigens and human fetal (i) erythrocyte antigen, and tumor-associated antigens such as sialyl Le". More recently, it has been shown that the lack of polylactosaminoglycan on the human erythrocyte anion transporter causes the glycoprotein to aggregate, resulting in

abnormal membrane structures in a congenital dyserythropoietic anemia-type II [9].

	NCBI Accession code	Lengh (aa)
Mus musculus	gi|13905006|gb|AAH06785.1|	189
Bos Taurus	gi|115497212|ref|NP_001068592.1|	409
Homo sapiens	gi|39645231|gb|AAH07845.2|	248
Rattus norvegicus	gi|6981144|ref|NP_036989.1|	407
Felis catus	gi|156447859|gb|ABU63691.1|	179
Gallus gallus	gi|45384206|ref|NP_990614.1|	414
Xenopus laevis	gi|147902288|ref|NP_001087042.1|	417
Macaca mulatta	gi|109121337|ref|XP_001087801.1|	416
Marmota monax	gi|121044661|gb|ABM46909.1|	322
Cricetulus griseus	gi|1346461|sp|P49129.1	407
Canis lupus familiaris	gi|73989504|ref|XP_534193.2|	413
Sus scrofa	gi|58332862|ref|NP_001011507.1|	413
Pan troglodytes	gi|114650748|ref|XP_001144542.1|	375
Equus caballus	gi|149730523|ref|XP_001495702.1|	400
Monodelphis domestica	gi|126337411|ref|XP_001374132.1	422

Table 1: LAMP-1 sequences with their length and NCBI accession code

2. MATERIALS AND METHOD

In order to search Lysosomal associated membrane proteins (LAMPs)family members we performed *BLAST* [1] by using blastp program in the protein database at *NCBI* [22].*Homo sapiens* LAMP-1 proteins' *gi|39645231|gb|AAH07845.2|* amino acid sequence was selected as query. From the hits 15 sequences, each from different species were selected for further studies. All the sequences were taken in *FASTA* format. The sequences were examined individually and aligned using *CLUSTALW* [11]. *Bioedit version 7.0.9.0* [10] was used for manual editing and analysis of sequences. *Kyte J and Doolittle* [13] method was used to plot hydrophobicity profile. Entropy is then calculated as:

$$H(l) = -\Sigma f(b,l)\ln(f(b,l))$$

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. As an alignment improves in quality, therefore, the entropy at each position (especially conserved regions) should decrease, which gives a measure of uncertainty at each position relative to other positions. Maximum total uncertainty will be defined by the maximum number of different characters found in a column.A window of defined size that was 13 is moved along a sequence, the hydropathy scores are summed along the window, and the average (the sum divided by the window size) is taken for each position in the sequence. The mean hydrophobicity value was plotted for the middle residue of the window. Eisenberg et. al. method [7] 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, θ =100 degrees.

$$\mu H = \{[Hn\sin(\delta n)]\text{^}2+[Hn\cos(\delta n)]\text{^}2\}$$

Where μH 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.

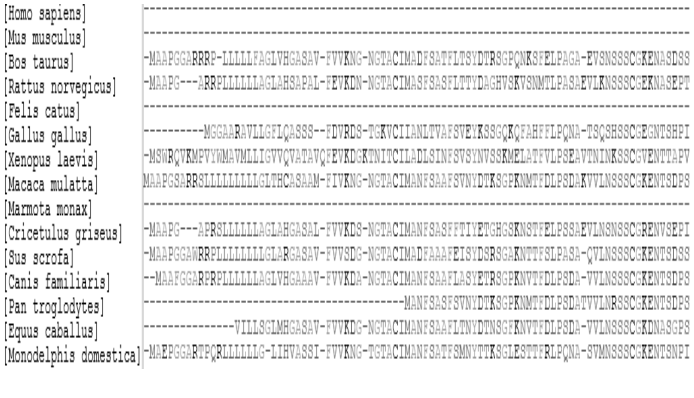
For a conserved region search within the multiple aligned sequences minimum segment length was set to 15 residues, maximum average entropy was set to be 2.0 and the gaps were limited to 2 per segment. Multiple sequence alignment, phylogenetic and molecular evolutionary analyses were conducted using *MEGA version 4* [19]. For pair wise and multiple alignments gap open penalty was -7 and gap extension penalty was -1. *BLOSUM* weight matrix was used for substitution scoring [2]. Hydrophilic gap penalties were used to increase the chances of a gap within a run (5 or more residues) of hydrophilic amino acids; these are likely to be loop or random coil regions where gaps are more common. The multiple alignmets of sequences of LAMP-1 proteins were used to create phylogenetic trees. The evolutionary history was inferred using the *Neighbour-Joining method* [17]. All the characters were given equal weights. The bootstrap consensus tree inferred from 20000 replicates [8] was taken to represent the evolutionary history of the taxa analyzed [8]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates were collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap tests (20000 replicates) are shown next to the branches [8]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the *phylogenetic tree*. The evolutionary distances were computed using the *poisson correction method* [23] and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). There were a total

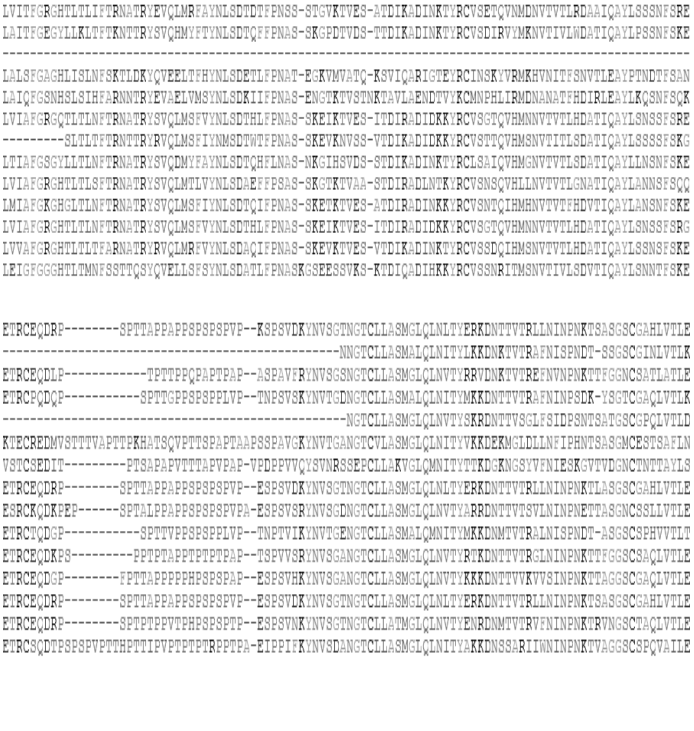
of 176 positions in the LAMP-1 proteins' final dataset. Phylogenetic analyses were conducted in *MEGA4* [9].

3. RESULT AND DISCUSSION

Multiple sequence alignment

The Multiple alignment of LAMP-1 proteins (Figure 1) resulted into an alignment having 429 positions. By statistical analysis of multiple aligned sequences it was observed that alanine, glycine, leucine, asparagine, serine, threonine and valine are the most frequently present amino acids with frequency percentage of 8.64, 6.22, 9.85, 7.00, 10.22, 8.45 and 7.71 respectively. While within conserved sites alanine, cystein, phenylalanine, glycine, leucine, asparagine, serine and valine are the most frequently occuring amino acids with frequency percentage of 10.70, 8.10, 8.72, 7.34, 12.00, 6.35, 6.96 and 6.27 respectively. The multiple aligned sequence of LAMP-1 protein was found with No. of conserved sites=99, No. of parsimony informative sites= 207, No. of singleton sites= 116 and no. of variables 328.





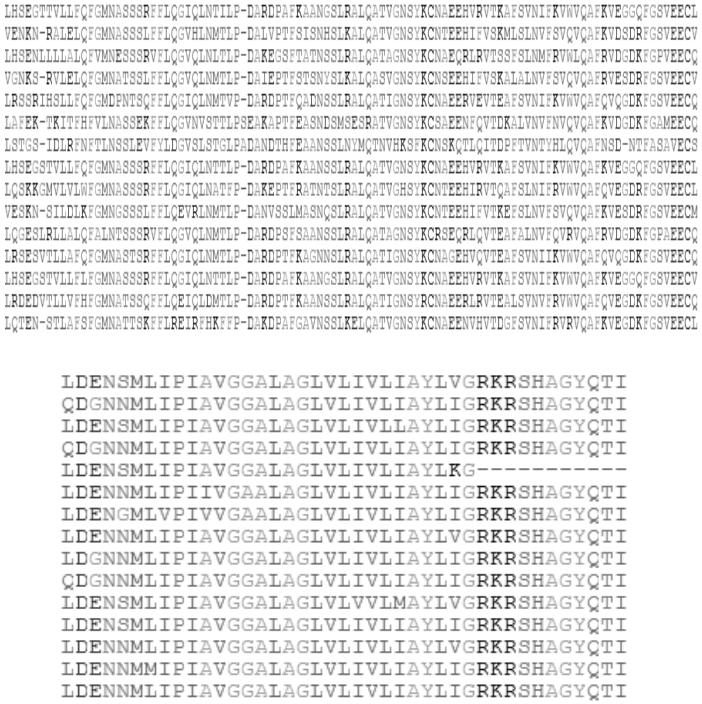


Figure1:. Multipe sequence alignment of LAMP-1 proteins.

Conserved region search

A conserved region search resulted into six regions from position 241 to 257 (segment length = 17), 259 to 276 (segment length = 18), 278 to 295(segment length = 18), 297 to 324 (segment length =28), 326 to 379 (segment length = 54) and 381 to 420 (segment length= 40) with an average entropy of 0.2152, 0.8244, 0.9500, 0.7522, 0.6430 and 0.2481 respectively as shown in figure 2. This conservation has already been upheld by minimal entropy shown by positions 240 to 420.

Region 1: Position 241 to 257
Segment Length: 17
Average entropy (Hx): 0.2152
Consensus:

241 NGTCLLASMGLQLNITY 257

Region 2: Position 259 to 276
Segment Lengths: 18
Average entropy (Hx): 0.8244
Consensus:

259 KKDNTTVTRLLNINPNKT 276

RKDNTTVTRLLNINPNKT
KKDNKTVTRAFNISPNDT
RVDNKTVTREFNVNPNKT
KKDNTTVTRAFNINPSDK
KRDNTTVSGLFSIDPSNT
KKDEKMGLDLLNFIPHNT
TKDGKNGSYVFNIESKGV
RKDNTTVTRLLNINPNKT
RRDNTTVTSVLNINPNET
KKDNMTVTRALNISPNDT
TKDNTTVTRGLNINPNKT
KKDNTTVVKVVSINPNKT
RKDNTTVTRLLNINPNKT
NRDNMTVTRVFNINPNKT
KKDNSSARIIWNINPNKT

Region 3: Position 278 to 295
Segment Length: 18
Average entropy (Hx): 0.9500

Consensus:

278 ASGSCGAQLVTLELHSEN 295

ASGSCGAHLVTLELHSEG
SSGSCGINLVTLKVENKN
FGGNCSATLATLELHSEN
YSGTCGAQLVTLKVGNKS
ATGSCGPQLVTLDLRSSR
ASGMCESTSAFLNLAFEK
VDGNCTNTTAYLSLSTGS
ASGSCGAHLVTLELHSEG
ASGNCSSLLVTLELQSKK
ASGSCSPHVVTLTVESKN
FGGSCSAQLVTLELQGES
AGGSCGAQLVTLELRSES
ASGSCGAHLVTLELHSEG
VNGSCTAQLVTLELRDED
AGGSCSPQVAILELQTEN

Region 4: Position 297 to 324
Segment Length: 28
Average entropy (Hx): 0.7522

Consensus:

297TVLAFQFGMNASSSRFFLQGIQLNMTLP 324

TVLLFQFGMNASSSRFFLQGIQLNTILP
RALELQFGMNASSSLFFLQGVHLNMTLP
LLLALQFVMNESSSRVFLQGVQLNLTLP
RVLELQFGMNATSSLFFLQGVQLNMTLP
HSLLFQFGMDPNTSQFFLQGIQLNMTVP
TKITFHFVLNASSEKFFLQGVNVSTTLP
IDLRFNFTLNSSLEVFYLDGVSLSTGLP
TVLLFQFGMNASSSRFFLQGIQLNTTLP
MVLVLWFGMNASSSRFFLQGIQLNATFP
SILDLKFGMNGSSSLFFLQEVRLNMTLP
RLLALQFALNTSSSRVFLQGVQLNMTLP
TLLAFQFGMNASTSRFFLQGIQLNMTLP
TVLLFLFGMNASSSRFFLQGIQLNTTLP
TLLVFHFGMNATSSQFFLQEIQLDMTLP
STLAFSFGMNATTSKFFLREIRFHKFFP

Region 5: Position 326 to 379
Segment Length: 54
Average entropy (Hx): 0.6430

Consensus:

326 DARDPTFKAANSSLRALQATVGNSYKCNAEEHVRVTKAFSVNIFKVWVQAFKVE 379

DARDPAFKAANGSLRALQATVGNSYKCNAEEHVRVTKAFSVNIFKVWVQAFKVE
DALVPTFSISNHSLKALQATVGNSYKCNTEEHIFVSKMLSLNVFSVQVQAFKVD
DAKEGSFTATNSSLRALQATAGNSYKCNAEQRLRVTSSFSLNMFRVWLQAFRVD
DAIEPTFSTSNYSLKALQASVGNSYKCNSEEHIFVSKALALNVFSVQVQAFRVE
DARDPTFQADNSSLRALQATIGNSYKCNAEERVEVTEAFSVNIFKVWVQAFQVQ
EAKAPTFEASNDSMSESRATVGNSYKCSAEENFQVTDKALVNVFNVQVQAFKVD
DANDTHFEAANSSLNYMQTNVHKSFKCNSKQTLQITDPFTVNTYHLQVQAFNSD
DARDPAFKAANSSLRALQATVGNSYKCNAEEHVRVTKAFSVNIFKVWVQAFKVE
DAKEPTFRATNTSLRALQATVGHSYKCNTEEHIRVTQAFSLNIFRVWVQAFQVE
DANVSSLMASNQSLRALQATVGNSYKCNTEEHIFVTKEFSLNVFSVQVQAFKVE
DARDPSFSAANSSLRALQATAGNSYKCRSEQRLQVTEAFALNVFQVRVQAFRVD
DARDPTFKAGNNSLRALQATIGNSYKCNAGEHVQVTEAFSVNIIKVWVQAFQVQ
DARDPAFKAANGSLRALQATVGNSYKCNAEEHVRVTKAFSVNIFKVWVQAFKVE
DARDPTFKAANSSLRALQATIGNSYRCNAEERLRVTEALSVNVFRVWVQAFQVE
DAKDPAFGAVNSSLKELQATVGNSYKCNAEENVHVTDGFSVNIFRVRVQAFKVE

Region 6: Position 381 to 420
Segment Length: 40
Average entropy (Hx): 0.2481

Consensus:

381DKFGSVEECQLDENNMLIPIAVGGALAGLVLIVLIAYLIG 420

GQFGSVEECLLDENSMLIPIAVGGALAGLVLIVLIAYLVG
DRFGSVEECVQDGNNMLIPIAVGGALAGLVLIVLIAYLIG
DKFGPVEECQLDENSMLIPIAVGGALAGLVLIVLLAYLIG
DRFGSVEECVQDGNNMLIPIAVGGALAGLVLIVLIAYLIG
DKFGSVEECQLDENSMLIPIAVGGALAGLVLIVLIAYLKG
DKFGAMEECQLDENNMLIPIIVGAALAGLVLIVLIAYLIG
NTFASAVECSLDENGMLVPIVVGAALAGLVLIVLIAYLIG
GQFGSVEECLLDENNMLIPIAVGGALAGLVLIVLIAYLVG
DRFGSVEECLLDGNNMLIPIAVGGALAGLVLIVLIAYLIG
DRFGSVEECMQDGNNMLIPIAVGGALAGLVLIVLIAYLIG
DKFGPAEECQLDENSMLIPIAVGGALAGLVLVVLMAYLVG
DKFGSVEECQLDENSMLIPIAVGGALAGLVLIVLIAYLIG
GQFGSVEECVLDENNMLIPIAVGGALAGLVLIVLIAYLVG
DKFGSVEECQLDENNMMIPIAVGGALAGLVLIVLIAYLIG
DKFGSVEECLLDENNMLIPIAVGGALAGLVLIVLIAYLIG

Figure2: Six conserved domains obtained with their length.

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 (Fig. 3). The plot shows that entropy rarely touches a scale of two, showing minimal entropy at several positions from position 245 to position 420 where entropy rarely crosses a scale of one, which is a sign of better alignment in the region. Any position before 240 doesn't show much conservedness.

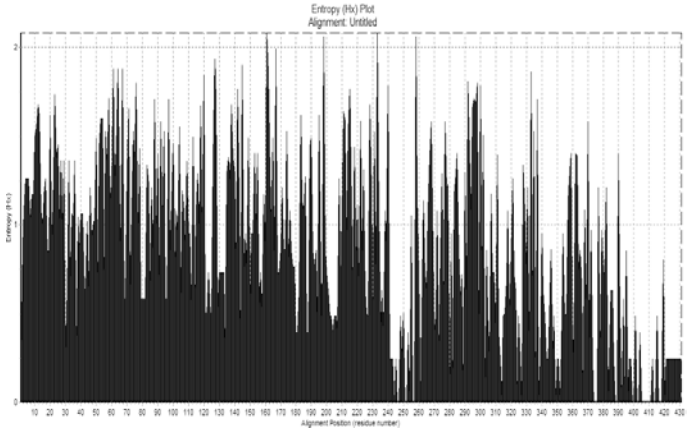


Figure3: Entropy (Hx) Plot

Hydrophobicity profile and hydrophobic moment

A hydrophobicity profile plot shows that mean hydrophobicity of the protein for most of the positions is in all the species is around zero, occassionaly it turns to be positive or negative (Fig.4). N-terminal domain and C-terminal domains are more hydrophobic. Maximum hydrophobicity is observed from positions 1 to 55 and between 390 to 420 positions in *Canis lupus familiaris*. *Macaca mulatta* and *Monodelphis domestica* exhibits high jump in hydrophobicity from position 10 to 40 and 400 to 420 respectively. 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.

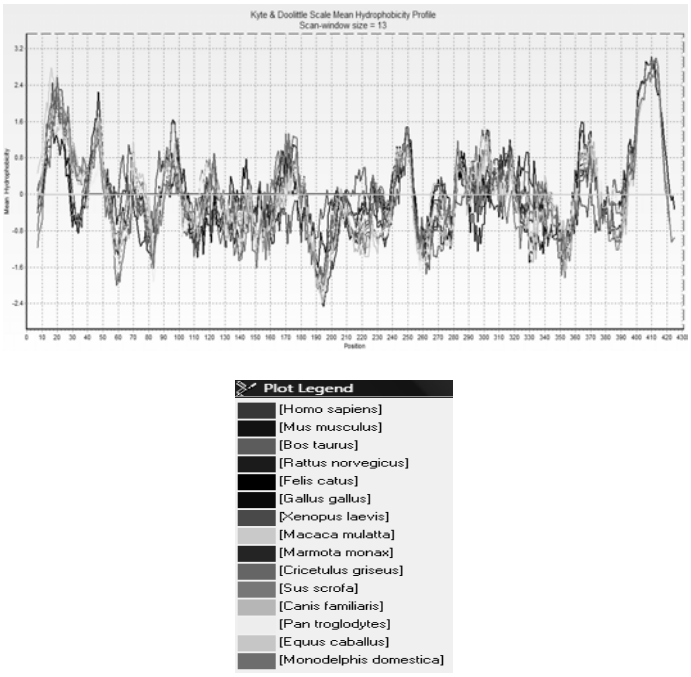


Figure4: Kyte and Dolittle scale mean hydrophobicity profile plot

Phylogeney

The phylogenetic tree constructed by using Neighbour –joining method (Fig. 5-6) shows different organisms on tree nodes branched on the basis of their LAMP-1 proteins. *Xenopus tropicalis* makes a totally diverged branch from the main tree among the selected proteins. Node for Mammalia is supported by lower bootstrap values i.e.99% while the node for primates(*Homo sapiens,Pan troglodytes,Macaca mulatta*) is supported by very high bootstrap value i.e. 100%. This tree gives an idea about the evolutionary order of LAMP-1 proteins. This phylogeny does not seem to be completely consistent with the current view of taxonomy perhaps due to use of a specific protein rather than complete genomes.

4. CONCLUSION

This study presents the first comparative genomic study and evolutionary analysis of the LAMP-1 proteins across family of organisms with special reference to mammals. The study established an overall framework of information for the family of LAMP-1 proteins, which may facilitate and stimulate the study of this gene family across all organisms.

5. FUTURE SCOPE

The evolutionary account of the Lysosomal membrane proteins will help us to reveal the some unexposed aspects of the many metabolic diseases linked to the lysosomal membrane proteins like Lysosomal storage diseases, I-cell disease etc. This work may facilitate to design the drug to over come these diseases.

6. ACKNOWLEDGEMENT

MD is thankful to DST for a Project Junior Research fellowship. The work has been supported by a DBT-BIF Grant to DKG under its BTISNet scheme and DST Project under its NanoScience and Technology (Nanomission).

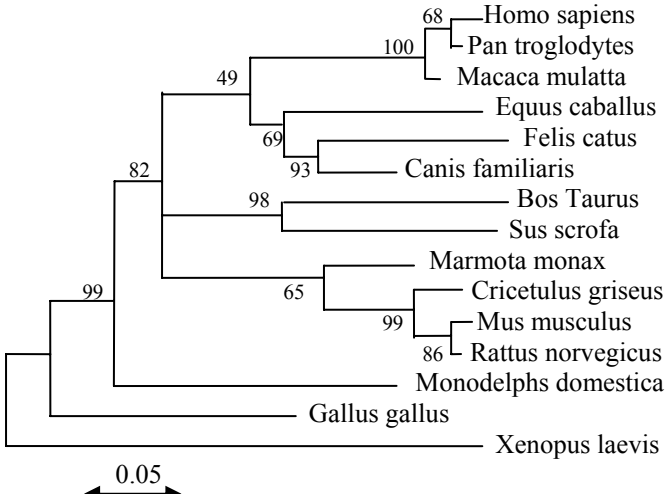


Figure5: Bootstrap consensus phylogenetic tree of LAMP-1 proteins created by Neighbour- joining method showing bootstrap support values on the nodes.

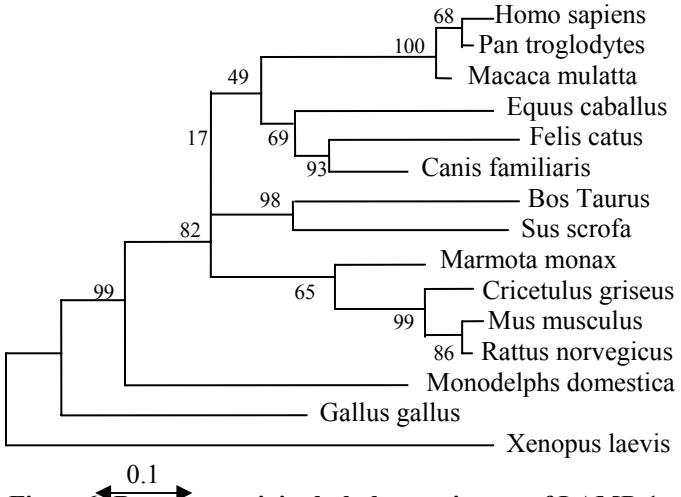


Figure6: Bootstrap original phylogenetic tree of LAMP-1 proteins created by Neighbour- joining method showing bootstrap support values on the nodes.

REFERENCES

[1]. Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. **25**:3389-3402, 1997.

[2]. Altschul S.F. and Gish G., Local alignment statistics Methods Enzymol. **266**:460-480 1996.

[3]. Bainton DF. The discovery of lysosomes. J Cell Biol ;91:66s–76s,1981.

[4]. Bernd Schroder, Christian Wrocklage, Cuiping Pan, Ralf Jager, Bernd Kosters, Helmut Schafer, Hans-Peter Elsasser,Matthias Mann and Andrej Hasilik, Integra and Associated Lysosomal Membrane Proteins, Journal compilation,1676-1686 , 2007.

[5]. Chi T. Hua, John J. Hopwood, Sven R. Carlsson, Ray J. Harris, and Peter J. Meikle, Evaluation of the lysosome-associated membrane protein LAMP-2 as a marker for Lysosomal storage disorders, Clinical Chemistry 44:10, 2094–2102, 1998.

[6]. Corinne Sagne, Cendra Agulhon, Philippe Ravassard et al Identification and characterization of Lysosomal transporter for small neural amino acids PNAS 98: 7206–721, 2001.

[7]. Eisenberg D. E. Schwarz, M. Komaromy and R.Wall. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J. Mol. Biol. 179(1):125-42, 1984 .

[8]. Felsenstein J ,Confidence limits on phylogenies: An approach using the bootstrap. Evolution **39**:783-791, 1985.

[9]. Fukuda, M. N., Dell, A., and Scartezzini, P., J. Biol. Chem. 262,7195-7206, 1987 .

[10]. Hall, T.A., BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41:95-98, 1999.

[11]. Higgins D., Thompson J., Gibson T. Thompson J. D., Higgins D. G., Gibson T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673-4680, 1994.

[12]. Jonathan N. Glickman and Stuart Kornfeld: Mannose 6-Phosphate-independent Targeting of Lysosomal Enzymes in I-Cell Disease B Lymphoblasts The Journal of Cell Biology, 123, 99-108, 1993.

[13]. Kyte J and Doolittle RF: A simple method for displaying the hydropathic character of a protien. J Mol Biol 157:105, 1982.

[14]. Liu H, Sadygov RG, Yates JR III. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal. Chem;76:4193–420, 2004.

[15]. Mancini GM, Havelaar AC, Verheijen FW. Lysosomal transport disorders. J Inherit Metab Dis;23:278–292, 2000.

[16]. Richard D. Bagshaw, Don J. Mahuran, and John W. Callahan: A Proteomic Analysis of Lysosomal Integral Membrane Proteins Reveals the Diverse Composition of the Organelle, The American Society for Biochemistry and Molecular Biology, Molecular & Cellular Proteomics 4.2, 133-143, 2005.

[17]. Saitou N & Nei M (**1987**) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution **4**:406-425, 1987.

[18]. Sven R. Carlsson, Jurgen Rothll, Friedrich Pillerz and Minoru Fukuda: Isolation and Characterization of Human Lysosomal Membrane Glycoproteins, h-lamp- 1 and h-lamp-2 ,The Journal of Biological Chemistryl. 263, 18911-18919, 1988.

[19]. Tamura K, Dudley J, Nei M & Kumar S MEGA4: Molecular Evolutionary Genetics analysis (MEGA) software version 4.0. Molecular Biology and Evolution **24**:1596-1599, 2007.

[20]. Town, M., Jean, G., Cherqui, S., Attard, M., Forestier, L., Whitmore, S. A., Callen, D. F., Gribouval, O., Broyer, M., Bates, G. P., et al. Nat. Genet. 18, 319–324, 1998.

[21]. Verheijen, F.W.,Verbeek, E.,Aula,N., Beerens, C. E.,Havelaar,A. C., Joosse, M., Peltonen, L., Aula, P., Galjaard, H., van der Spek, P. J. & Mancini, G. M. Nat. Genet. 23, 462–465, 1999.

[22]. www.ncbi.nlm.nih.gov/entrez (National Centre for Biotechnology Information).

[23]. Zuckerkandl E & Pauling L, Evolutionary divergence and convergence in proteins, pp. 97-166 in Evolving Genes and Proteins, edited by V. Bryson and H.J. Vogel. Academic Press, New York, 1965.