(Review Article)

IJARPB

International Journal of Advanced Research in Pharmaceutical & Bio Sciences
(A peer reviewed International journal for Pharmaceutical and Bio Sciences)
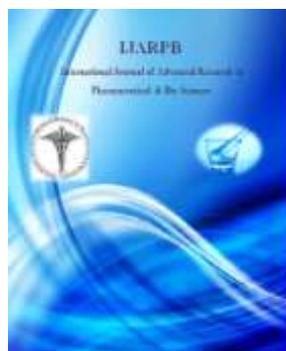
Bioinformatics: Phylo-genetic comparison of *Kappa Casein* Protein Family

Pareek; K[*1], Shah; S[1] and Asha Arora[**2]

[1] School of life Science, Pacific University, Udaipur

[2] Head, Deptt. of Botany and Biotechnology, B.N.P.G.Girls College,Udaipur

Corresponding Author

Pareek; K

School of life Science, Pacific University, Udaipur

E mail: kbiopareek@gmail.com

ABSTRACT

K-Casein; is a mammalian milk protein belong to phosphoproteins family and involved in a number of important physiological processes. We are analyzed 42 members of Kappa-casein reported in uniport kb database. Out of which maximum length of protein sequence 234Amino Acid (P19442) and Minim length 122 AA (09-members) are long. Out of candidate protein family (P02669) consists 190AA and Cleaved into 5 chains (Casoxin-C, Casoxin-6, Casoxin-A, Casoxin-B, Casoplatelin), and rest of all are cleaved into 2 chain except (Q29150AA) which is non Cleaved chain. Maximum frequency (122AA length, 9 times) and minimum are (13 different length, only one time are notice. There is also remark that cleaving frequency amino acids site positions (58-59) were maximum 9 times in candidate protein family, but ranging differences are maximum 17 in between the length of 153.The multiple sequence alignment using Clustalx (version 2.0.11) with Pearson method alignment score 464342 obtain in that family and Phenetic (Distance Method) using Neighbor Joining algorithm (NJ Plot) for build tree. Calculating Bootstrap replicates to Phylogeny analysis, the distance scale (0.05), seven major clusters are identified and thirteen terminal common ancestors found in 42 nodes.

Keyword: cleaving   frequency,   multiple   sequence   alignment,   Neighbor   Joining algorithm, distance scale

(Review Article)

INTRODUCTION:

Casein is milk protein secreted by mammary gland cells. Family of phosphoprotein is subdivided into four main groups: αS1-casein, αS2-casein, β-casein, and κ-casein. In Cattles caseins gene comprehends a 200 kb fragment in of chromosome 6, arranged in tandem in the following order: α-s1, β, α-s2 and κ.[5][7]. Specifically, the κ-casein gene comprehends a 13 kb sequence divided into 5 exon [1]. For several buffalo breeds; the genetic variability in the κ-CN locus has been reported each with a different allelic frequency based on genetic diversity among breeds. Various allelic variants have been described for κ-CN gene in different cattle breeds, which include A, B, C, E, F, G, H, I and AI [2] Among these, variants A and B are most commonly found and variant B is predominantly concerned with processing properties of milk and has better lactodynographic properties [6].

The casein protein is divided into two groups, the calcium-sensitive and the non-calcium-sensitive. The κ-casein is insensitive to calcium and α(s1)- and α(s2)-caseins and β-casein are calcium-sensitive. Casein protein chemical and physical properties (Molecular weight 19,000 KDa.; 169 residues, 20 prolines), and very resistant to calcium precipitation. Rennet cleavage at the Phe105-Met106 bond eliminates the stabilizing ability, leaving a hydrophobic portion, para-kappa-casein, and a hydrophilic portion called kappa-casein glycomacropeptide (GMP), or more accurately, caseinomacropeptide (CMP)

Table1: Nomenclature History of Kappa Casein Protein

Symbol	Name	Reference	Event	Date
Csnk	casein kappa	J:34481	assigned	09/23/96
AW208918	expressed sequence AW208918	J:68900	assigned	08/14/01
AW208918	expressed sequence AW208918	J:75000	withdrawn, = Csnk	03/26/02
Csnk	casein kappa	J:85834	withdrawn, = Csn3	08/10/06
Csn3	casein kappa	J:85834	assigned	08/10/06

There are several models that account for the special conformation of casein in the micelles, the most recent model [3] proposes a double link among the caseins for gelling to take place. The UCSF Chimera Alpha version 1.3 ,University of California for protein structure prediction shows the Kappa casein protein surface structure and Secondary structure with Helix and Strand Structure.
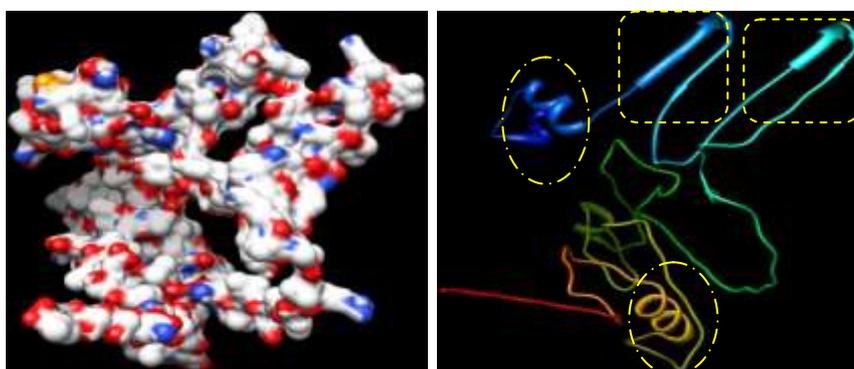

Figure: 1-a Surface Model of Kappa Casein and Figure: 1-b Secondary Structure (i) Oval Shape is Helix form and (ii) Rectangle Shape is Strand form

(Review Article)

METHODOLOGY

Sequences

An exhaustive search was made in GenBank and EMBL databases [12] for Kappa casein milk protein Family. This search was optimised by protein knowledge base UniportKb Data base were protein sequences obtained and then filtered by removing partial and redundant sequences from the population. Complete protein representations by strain were included; our final working population consisted of 42complete protein sequences [13]

Analyzing trace residues

The MSA of protein sequences obtained from Clustalx [4],[14] , including the Family of candidate protein sequences, input file for MSA are used in Fasta format obtain from Uniportkb database. The phylogenetic tree

Multiple Sequence Alignment and phylogenetic reconstruction

General help for CLUSTAL X (2.0) [4] [14] was used for protein multiple Sequence Alignments (MSA)[15] using the 42 previously collected sequences with standard parameters and NJ [10] methods were used for phylogenetic reconstruction and the *p-distance* method was used for distance analysis [8]. Base statistical robustness was ensured by using 1000 Bootstrap repeats. The complete process was developed by Nj plot tools [10] for build tree.
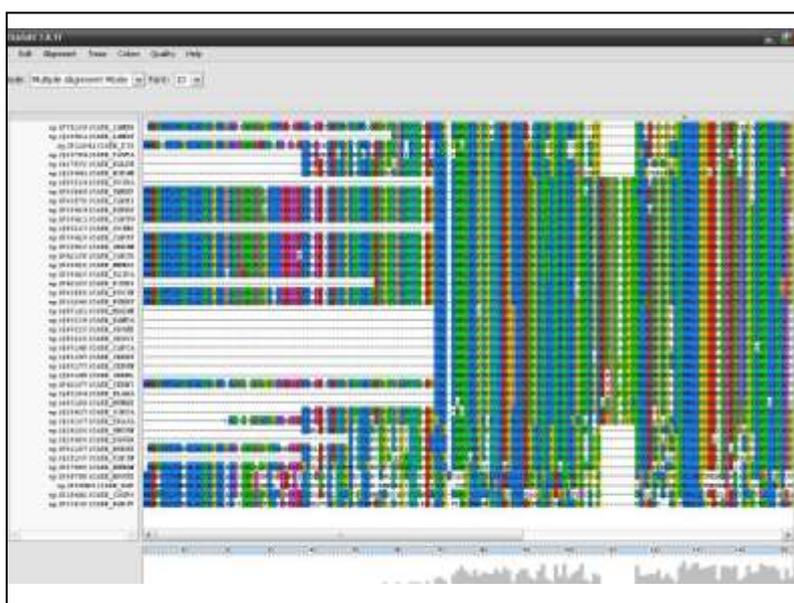

Figure2: MSA of Kappa Casein Protein Family in Clustal x 2.00 software package and alignment was done employing conserved sequence.

Was inferred from the multiple sequence alignments of candidate family in proper sequence datasets, separately using NJ Plot package [10]. In MSA were starting points for pair wise sequence identity was calculated for all the combinations. A Cladogram was created from the multiple sequence alignment of these 42 Members through the NJ Plot package using the distance matrix. The

bootstrap value (more than 10000) and the sequence identity (Default) parameters were considered for clustering the members. In some clusters, the members are clustered together even if they satisfy one parameter, either the bootstrap value or the sequence identity. Branches of the Cladogram were clubbed together if the bootstrap value was more than 50. The bootstrapping of the

(Review Article)

multiple sequence alignment was performed 1000 times for the 'alignment dataset' through SEQBOOT. Protein distances were calculated employing Clustal x2.0 package. The trees were calculated using the rooted Neighbor-Joining (NJ) method [10] on distance matrices employing NEIGHBOR from the Clustalx package. These trees were analyzed employing of Clustalx package to derive a consensus tree. A rooted tree was plotted using Both TREEVIEW 1.6.60 [16] and NJ Plot software package. Sequences with more than 50% bootstrap support values were confirmed and grouped.

RESULT AND DISCUSSION:

Sequence Length	No. of Frequency	Swiss-Prot ID	Cleavage Position	Chain Cleavage
122	9	Q95147,Q95149,Q95177,Q95184,Q95191,Q95199,Q95225,Q95228,Q95239	58-59	02
123	1	Q95146	58-59	02
124	2	Q95227,Q95224	58-59,58-52	02
125	1	Q28451	59-60	02
135	1	Q28400	70-71	02
136	1	P42155	72-73	02
145	2	Q27952,Q28441	81-82	01,02
146	3	Q29135,Q28794,Q29150	80-81	02
153	1	Q28417	89-90	02
171	1	P29137	107-108	02
178	1	P04468	155-166	02
180	1	P33618	166-177	02
181	1	P06796	118-119	02
182	2	P79139,P07498	117-118	02
185	1	P82187	117-118	02
188	1	P11841	117-118	02
190	3	P02668,P11840,P42157	(1-21,22-190) 126-127	02
192	8	P02670,P42156,P50420,P50421,P50422,P50423,P02669,P50424	126-127	02
202	1	P50425	126-127	02
234	1	P19442	118-119	02

Table2: Sequence length Cleavage analysis of Kappa Casein Protein Family

(Review Article)

In Candidate protein family maximum length of protein sequence 234AA (P19442) and Minim length 122 Amino Acid (09-members) are long. Out of candidate protein family (281728) consists 190AA and Cleaved into 5 chains (Casoxin-C, Casoxin-6, Casoxin-A, Casoxin-B, Casoplatelin), and rest of all are cleaved into 2 chain except Uncia uncia (Q29150AA) which is non Cleaved chain. Maximum frequency (122AA length, 9 times) and minimum are (13 different lengths), only one time are notice. There is also remark that cleaving frequency of amino acids site positions (58-59) were maximum 9 times in candidate protein family, but ranging differences are maximum 17 in between the length of 153.

The multiple sequence alignment (Fig2) using Clustalx (version 2.0.11) with Pearson method alignment, blossom 62 matrix used to calculate distance for obtain score. Input fasta format file which is pre downloaded from uniportkb data base and save in default (uniprot-family-kappaa-casein-family.fasta) and obtain optimal output. The output file (uniprot-family-kappaa-casein-family.aln) saved in local hard disk and view in clustalx packeg and save as in postscript format.The result obtained score are 464642 and conservation of total alignment is less then 25% gaps.

Using crustalx package for build phylogeny tree used Calculating Bootstrap replicates to Phylogeny analysis, the distance scale (0.05) pre default parameter used and boot strap file uniprot-family-kappaa-casein-family.phb obtain and finally obtain complete tree. To view the result used NJ plot software package and Tree View 1.1 package [16]. Obtain result save in JPEG file format and also save in portable document format (*.pdf) in local disk.

The phylogenetic Protein tree developed for Kappa Casein 7 main clades using the NJ method [10]; tree topology was entirely similar to results obtained using the UPGMA algorithm [11] Fig. 3A. These were derived from NJ Plot Package in the Cladograme, protein directed tree [9] clearly showed an ancestral division, this being coherent with that observed in the Cladograme tree constructed in the Local Machine. The topology of the Kappa Casein trees multiple levels events which have defined the Dept of tree, resulting in 7 main clad in both methods, with strongly supported by bootstrap values. In figure 3a bootstrap value 0.5 to calculate relation between existing members and construct tree with multi level depth is found. The arrow indicate an origin from at least two common ancestors Out of 42 existing members there are 13pairs observed that common ancestors in different dept in tree, The length is0.041, sp(p33618), sp(19442) only one pair is most nearby(level 01) from root of tree. Total level are 13from the root. Table 3 show the 7 clad of Phylogenic tree. Out of 42 members 35member are clubbed in major clad rest are freeware in tree.
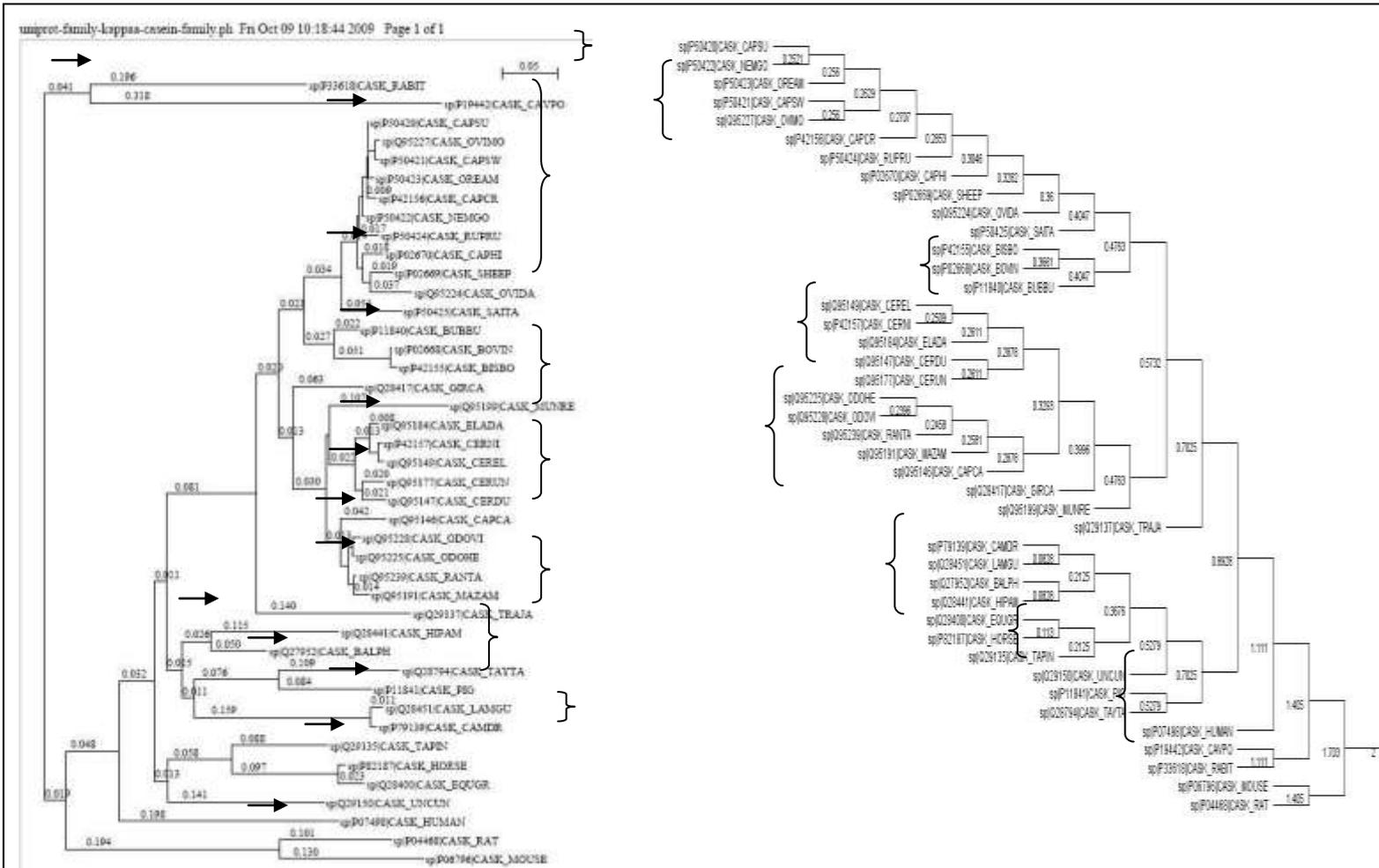
Fig3 (1) show the Cladograme calculated by NJ algorithm in Clustalx and picture generated in NJ plot software package marked define the major clades (2) calculated UPGMA Algorithm show the similarities between cluster

(Review Article)

Swiss-Prot ID	Clad	Member
P33618, P19442	01	02
P50420,Q95227,P50421,P50423,P42156,P50422, P50424,P02670, P02669, Q95224, P50425	02	11
Q95184, P42157, Q95149,Q95177, Q95147,	03	05
Q95146, Q95228, Q95225, Q952239, Q95191	04	04
Q28441, Q27952, Q28794, P11841, Q28451,P79139,	05	06
Q29135, P82187, Q28400, Q29150, P07498	06	05
P04468, P06796	07	02

Table 3 members of define 07 major clad of Phylogeny tree

Swiss-Prot ID	Clad	Common Name	Between Lenght
P33618, P19442	01	*Cask –Rabbit,Cask-Cavpo*	(0.041) 0.196,0.318
Q95227,P4421	02	*Cask-Ovimo, Cask-Capsw*	(0.014)
P02669,Q95224	02	*Cask-Sheep, Cask-Ovida*	(0.014) 0.019,0.037
P02668,P42155	NA	*Cask-Bivin, Cask Bisbo*	(0.051)
P42157,Q95149	03	*Cask-Cerni,Cask-Cerel*	(0.013)
Q95177,Q95147	03	*Cask-Cerun,Cask-Cerdu*	(0.022) 0.020,0.021
Q95228,Q95225	04	*Cask-Ovdovi,Cask-Odohe*	(0.013)
Q95239,Q95191	04	*Cask-Ranta, Cask-Mazam*	(0.014) 0.014
Q48441,Q27952	05	*Cask-Hipam, Cask-Balph*	(0.026) 0.115,0.050
Q28794,P11841	05	*Cask-Tayta, Cask-Pig*	(0.076) 0.109,0.084
Q28451,P79139	05	*Cask-Lamgu, Cask-Camdr*	(0.159) 0.011
P82187,Q28400	06	*Cask-Horse, Cask-Equgr*	(0.097)0.023
P04468, P06796	07	*Cask-Rat, Cask Mouse*	(0.194) 0.101,0.130

Table 4: 13 Arrow symbol are shown thirteen terminal common ancestors identified in 42 nodes. "()"value in parenthesis Define the common ancestors length from his upper root-node and other is own member length from near ancestors

(Review Article)

Table 4 shown the 13 pair of common ancestors identified in 42 nods in phylogeny tree. P02668 (*Cask_Bovin*) and P42155 (*Cask_Bisbo*) is only Pair that is not Clubbed in seven Cluster in Phylogeny tree, and this pair are closed relation in P11840 (*Cask_Bubbu*). See Fig-3(i), P33618 (*Cask_rabit*) and P19442 (*Cask_cavpo*) is more far evolutionary relationship to other family members.

CONCLUSION

The present study is a step forward towards a better understanding of the distribution of Kappa Casien Family in the Milk protein genera. The study has identified the 13 common ancestors on bases of phylogeny analysis. It is interesting to note that the relationships revealed by these studies are identical; we have also identified a few classes which will be highly useful with respect to their evolutionary analysis. This Protein Sequence studies is used for the sustainable research for carrying out genetic and phylogenetic studies.

Hopefully, this Study has revealed the importance and utility of bioinformatics resources that are valuable for current Proteomics research. The resources and tools reviewed in this paper are used mostly for Sequence analyses and Phylogeny analyses, which are essential for finding evolutionary relationship and genetics similarities.

REFERENCES

1. Alexander LJ, Stewart AJ, Machinlay AG, Kapelinskaya TV, as an aid for genetic improvement of dairy bulls and cows: A review. Animal Breeding Abstract 60:1-10

2. Soria,L.A.., G.m. Iglesias,M.U. Huguet and S.L. Mirande,2003. A PCR –RFLP test to detect allelic varients of the bovine kappa-casein gene. Anim Biotech., 14:1-5.

3. Horne, D. S. 1998. Casein interactions: Casting light on the black boxes, the structure in dairy products. Int. Dairy J. 8:171–177.

4. Larkin,M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G. (2007) Clustal W and Clustal X version 2.0. Bioinformatics, 23:2947-2948.

5. Eigel, W. N., J. E. Butler, C. A. Ernstrom, H. M. Farrell, Jr., V. R. Harwalkar, R. Jenness, and R. M. Whitney. 1984. Nomenclature of proteins of cow's milk: Fifth revision. J. Dairy Sci. 67: (1599– 1631).

6. Lin CY, Sabour MP and Lee AJ (1992) Direct typing of milk proteins

7. Lipids in Health and Disease, 2007 "Bovine milk in human nutrition – a review" Anna Haug, Arne T Høstmark and Odd M Harstad. Biomed central journal 6:25 (476-511).

8. Nei, M. and Kumar, S. (2000). Molecular Evolution and Phylogenetics. Oxford University Press, New York. of the bovine kappa-casein gene. Europe Journal Biochemical 178:395-401.

9. Penny, D., Hendy, M. D. & M. A. Steel. 1992. Progress with methods for constructing evolutionary trees. *Trends in Ecology and Evolution* **7**: 73-79.

(Review Article)

10. Saitou, N. and Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406-425.

11. Sokal, R. R. and Michener, C. D. (1958). A statistical method for evaluating systematic relationships. The University of Kansas Science Bulletin 28, 1409-1438.

12. Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007). MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596-1599.

13. www.uniportkb.family-kappa-casein family in UniProtKB.html

14. Thompson,J.D., Gibson,T.J., Plewniak,F., Jeanmougin,F. and Higgins,D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research, 25:4876-4882.

15. Humphrey, W., Dalke, A. and Schulten, K. (1996).- Visual Molecular Dynamice. J.Mol. Graph. 14, 33-38.

16. http://taxonomy.zoology.gla.ac.uk/rod/treeview/help/contents.html Tree view

OTHER SUPPORTING LINK

17. www.Informatics.jax.org
18. www.Informatics.jax.org)
19. www.ncbi.nlm.nih.gov
20. www.Informatics.jax.org