DOI : 10.35124/bca.2019.19.2.4425

Biochem. Cell. Arch. Vol. 19, No. 2, pp. 4425-4429, 2019 | www.connectjournals.com/bca | *ISSN 0972-5075*

CLASSIFICATION NUMBER OF ORGANISMS USING CLUSTER ANALYSIS OF THE PEPTIDE CHAINS MULTIPLE CHYMOTRYPSIN LACTATE DEHYDROGENASE

Ali A. Ibrahim[1], Mohammed Isam[2,3*] and Duaa Hammoud[4]

[1]Department of Mathematics, Al-Nahrainuniversity, Baghdad, Iraq.
[2]Department of Medical Microbiology, Koya University, Erbil, Iraq.
[3]Genome Center, Koya University, Erbil, Iraq.
[4]Department of Biotechnology, Al- Nahrain University, Baghdad, Iraq.
*e-mail: mohammed.isam@koyauniversity.org

(Received 24 March 2019, Revised 17 June 2019, Accepted 30 June 2019)

ABSTRACT : The current research aims to classify and to build a phylogenetic tree to 19 different organisms based on a series peptide multi-enzyme lactate dehydrogenase using the method of clustering analysis, one of the most important results of this research classify and distinguish organisms into six groups, each group including a number of different organisms are similar in among them. The idea behind of this research paper are building phylogeny tree of chymotrypsin from different species like (turtle, ostrich, dove, bee, koala, walrus, dolphins, whale, camel, monkey, mouse, squirrel, tiger, chicken) then comparing between the groups to get optimal score for which one of the observations or objects in every cluster square measure similar and therefore, the clusters square measure dissimilar to every alternative a result similar to multi chain peptide enzyme lactate dehydrogenase and organism differ in one group from the other groups.

Key words : Lactate dehydrogenase, phylogenetic, clustering analysis, NCBI.

INTRODUCTION

Chymotrypsin and trypsin are serine proteases with similarities in structure and sequence; in the other hand they are different in substrate specificity. And the previous experiments shows us the important role for the two loops that binding outside the pocket however its dominate the 2 enzymes specificity **(**Wenzhe Ma, 2005).

Chymotrypsin is aromatic residues shape like tyrosine, tryptophan and phenylalanine. When replacing two loops of trypsin L1 and L2 with loops of chymotrypsin and also replaced with mutation D189S, the activity of chymotrypsin increased in new protein around more than 1000- fold against mutant D189S (Chao Tang, 2005).

Chymotrypsin is an enzyme known as protein-digesting the pancreas secreted this enzyme. Play an important role by monitoring patients or people who suffering with pancreatic problem or dysfunction of pancreas. The pancreas is responsible to synthesized of chymotrypsin this process start from biosynthesis protein like a precursor known as chymotrypsinogen that is ambiguously inactive. The two enzymes (chymotrypsin and Trypsin) have high identity in their sequence and also they have high similarity in their tertiary structures (Fig. 1).

The goals behind this research are:

- Building phylogeny tree of chymotrypsin from different groups.
- Comparing between the groups to get optimal score for which one of the observations or objects in each cluster are similar.

MATERIALS AND METHODS

Sequences retrieval

The 30 full length amino acid sequences of chymotrypsin was collected (download) from Genebank databases (NCBI). http://www.ncbi.nlm.nih.gov

The thirty species are (cattle, buffalo, walrus, dolphins, whale, camel, monkey, mouse, squirrel, tiger, chicken, wolf, zebra, sheep, beetle, cockroach, falcon, mosquito, horse, parrot, turtle, ostrich, dove, bee, koala, octopus, spider, carb, goose and frog).

Bioinformatics tools and programming

The first step in this research, by using three of different software's and their packages :

1. Multi sequence alignment of individual profiles was

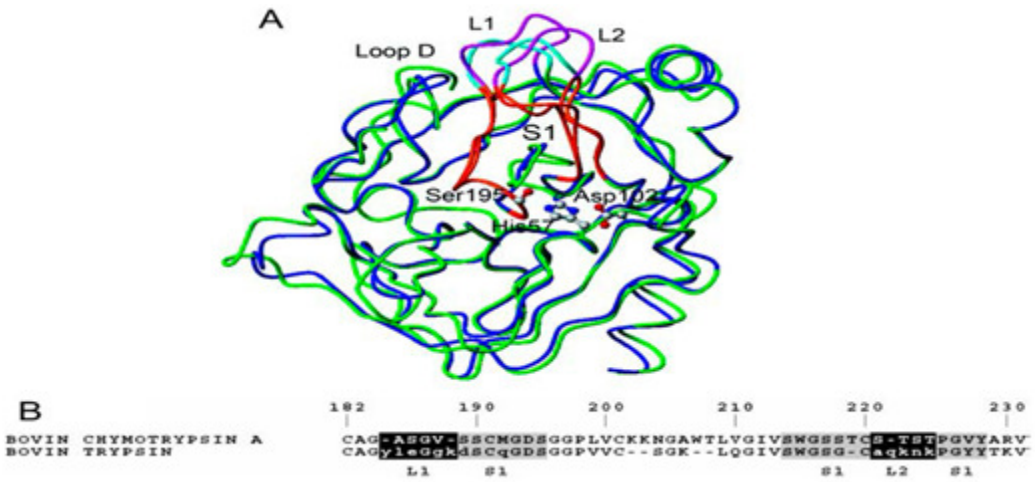

Fig. 1 : Tertiary structures of chymotrypsin and Trypsin.

Step 1. Distance matrix computation

Step 2. Start merge the closet two clusters

Step 3. Distance matrix update to retum the proximity between original clusters and the new cluster

Step 4. Go to step 2

Step 5. Until only one cluster remains

Fig. 2 : Agglomerative hierarchical clustering algorithm.

performed by using Cluster X.

2. Cluster analysis : Alignment of sequences (from the previous step) read it in dnadist PHYLIP as software used for convert data (our sequence of DNA) into distance matrix.
3. SPSS : The distance matrix read it in SPSS software for performance clustering analysis.

Hierarchical cluster analysis

Cluster analysis technique is one of the multivariate methods that used to find different patterns in a dataset by grouping as clusters. The hierarchical cluster is cluster techniques used for analysis, it takes a fraction a part of groups and classified knowledgeto clusters then these clusters are terribly shut just like others considering similarity inside cluster, in other hand is different or dissimilar between the clusters.

The hierarchical cluster method is taken into account a perfect, an ideal, good tool for analysis (data).

This analysis technique consists of (n) elements and for each element hasvariety of variables (p). According to this two basic approaches for creating hierarchical clustering are available:

Agglomerative: Initiation with points consider as individual cluster thenfor every step merges the nighest combine of clusters.

Divisive : Initiation with one of all-inclusive cluster then for each step divided to a cluster till only one of individual point singleton cluster remain. At this moment needs to decide which cluster to divide at each step and how to do the dividing.

Also by this method can display graphic way using a tree as a diagram known a dendrogram which displays both cluster sub-cluster and show uswhen the cluster incorporated or divide for sets of 2 dimensional points, these points were cluster exploitation the single-techniques which totallyvariations on a single-approach: ranging from individual points as cluster in turn merge the 2 nighest clusters till solely one cluster remains. This method is explain more in formally way (Fig. 2).

Form Fig. 2 can be explained as follows:

A. Calculating descent distance that decides and specifies closing degree between every two types of different elements according to the following equation:

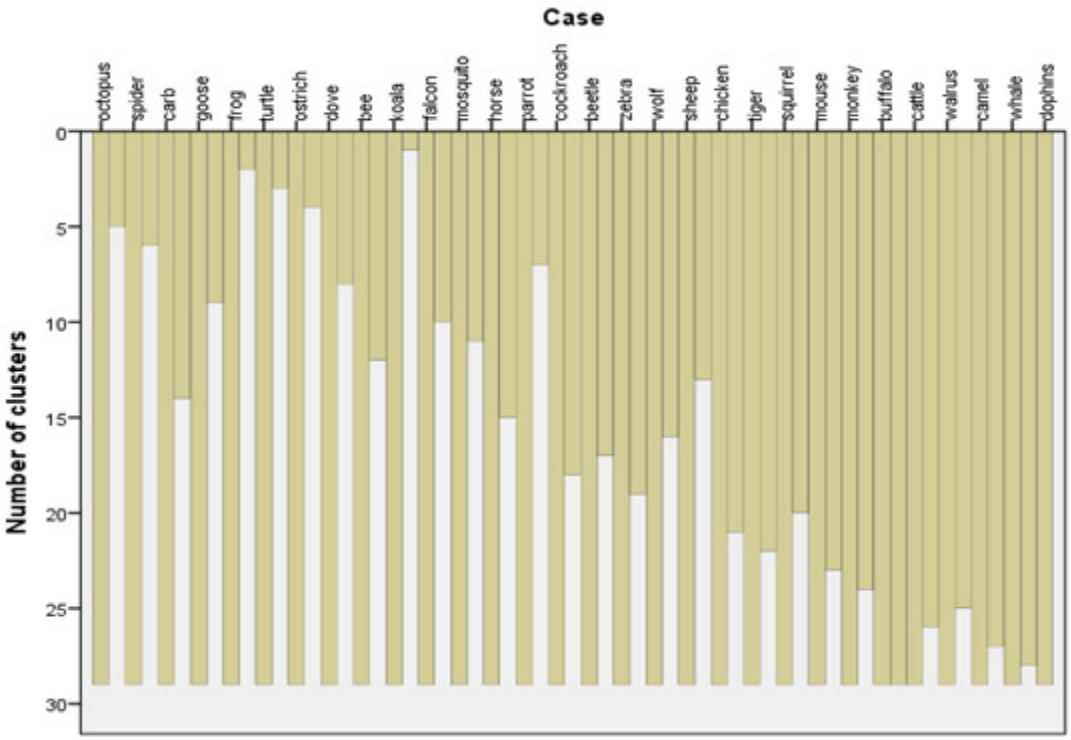

Fig. 3 : Clusters steps.

Table 1 : Distance of cluster steps.

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	20	23	.096	0	0	4
2	1	5	.151	0	0	3
3	1	2	.305	2	0	5
4	10	20	.432	0	1	5
5	1	10	.514	3	4	6
6	1	19	1.087	5	0	7
7	1	15	1.414	6	0	10
8	3	17	2.186	0	0	9
9	3	18	2.303	8	0	10
10	1	3	2.975	7	9	17
11	24	25	4.024	0	0	13
12	27	28	7.980	0	0	13
13	24	27	13.079	11	12	14
14	21	24	15.533	0	13	17
15	8	16	16.081	0	0	19
16	11	26	20.067	0	0	21
17	1	21	20.648	10	14	23
18	4	29	24.841	0	0	22
19	8	14	24.966	15	0	20
20	8	22	29.187	19	0	23
21	9	11	32.814	0	16	24
22	4	7	34.819	18	0	26
23	1	8	37.874	17	20	29
24	9	12	45.713	21	0	25
25	9	30	48.084	24	0	28
26	4	6	49.565	22	0	27
27	4	13	69.464	26	0	28
28	4	9	73.864	27	25	29
29	1	4	99.797	23	28	0

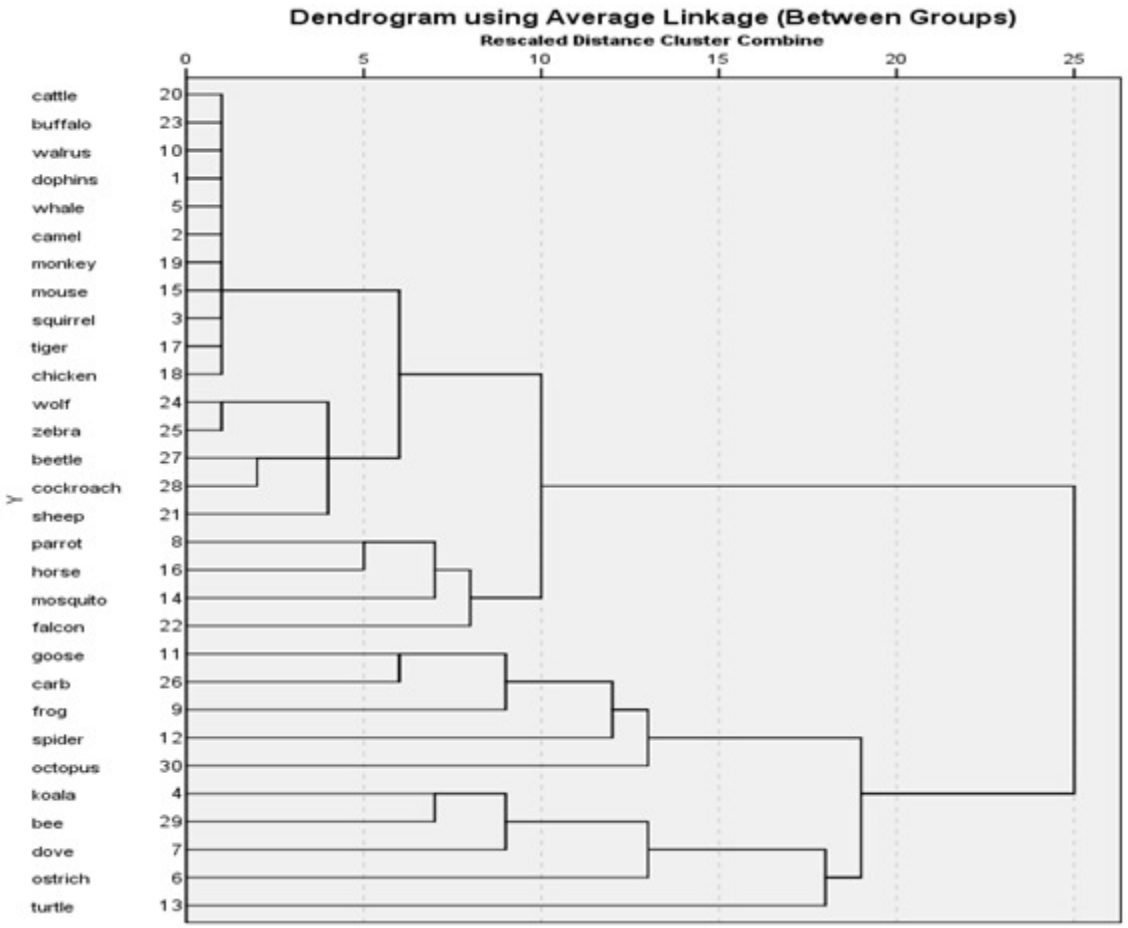

Fig. 4 : Dendrogram between tree using averages between groups.

$$d_{ij} = \sqrt{\left\{\sum_{k=1}^{p}\left(x_{ik} - x_{jk}\right)\right\}} \quad (1)$$

B. find the shortestdistance in matrix (d_{ij}) to create relationship between first element (i) and second element (j), x is variable of the value for individual J.

C. Containing of Agglomerate or grouping this depending on the possible shortest descent distance, till making connection of the last two groups or cluster to form the final groups (7, 9).

RESULTS AND DISCUSSION

After entering the DNA sequences of twenty one species into cluster – w software to align these sequences next step was to enter the aligned sequences into Dandiest PHYLIP software to get thedistance matrix (Table 1).

And last step was to enter the distance matrix into SPSS package to implement cluster analysis to build a phylogenetic tree (Figs. 3 and 4).

The results of cluster analysis from (**Fig. 3)**describe the process as possible clustering among the different races of organisms, were six groups. The first group contains 11 species Cattle, buffalo, walrus, dolphins, whale, camel, monkey, mouse, squirrel, tiger, chicken. While second group contains 3 species wolf, zebra, sheep. The third group contains 2 species Beetle, cockroach. The fourth group contains 4 species Falcon, mosquito, horse, and parrot. The fifth group contains 5 species turtle, ostrich, dove, bee, and koala. The last group (sixth group) contains 5 species Octopus, spider, carb, goose, and frog. These groups contains similar (objects) species (within groups), while each group contains dissimilar (objects) species with the other groups (between groups).

Recommendation

A- Increase the number of peptide chains and multiple different enzymes to determine the extent of convergence between genetic organisms.

B- Use factor analysis method to analysis the raw data.

C- Building a database containing a number of multi –peptide chains and enzyme for a variety of different organisms.

REFERENCES

Blow D M (1997) The tortuous story of Asp… His… Ser: structural analysis of alpha-chymotrypsin. *Trends Biochem. Sci*. **22**, 405–408.

Bryan F J Manly (1986) Multivariate statstical methods aprlmer. University of Otego, Newzeland, Chapman and Hall.

Elustondo P A, White A E, Hughes M E, Brebner K, Pavlov E and Kane D A (2013) Physical and functional association of lactate dehydrogenase (LDH) with skeletal muscle mitochondria. *J. Biol. Chem*. **288**, 25309–25317. doi: 10.1074/jbc.M113.476648

Fieller N (2001) Further Multivariate Analysis :Working Notes, NRJF, University of Sheffield.

Hastie Trevor, Tibshirani Robert and Friedman Jerome (2009) "14.3.12 Hierarchical clustering". *The Elements of Statistical Learning* (PDF) (2nd Ed.). New York: Springer. pp. 520–528.

Joseph F Hair Jr (1987) Multivarialate Data Analysis with Readings. Macmillan Polishing Company, New York.

Kaufman L and Rousseeuw P J (1990) *Finding Groups in Data: An Introduction to Cluster Analysis* (1 ed.). New York: John Wiley. ISBN 0-471-87876-6.

Maurice Kendall Sc D F B A (1987) Multivariate Analysis. Charles Griffin and company LTD, London and High Wycombe.

National Center for Biotechnology Information.http:// www.ncbi.nlm.nih.gov

Petsko Gregory and Ringe Dagmar (2009) *Protein Structure and Function*. Oxford: Oxford University Press. pp. 78–79.

Salvatore Passarella, Gianluca Paventi and Roberto Pizzuto (2014) The mitochondrial L-lactate dehydrogenase affair. *Frontiers in Neuroscience.*

Zenko B (2007) Learning predicative clustering rules. *Ph D thesis*, faculty of computer and information sciences, University of Ljubljana.