ISSN No. 0976-5697

Volume 5, No. 5, May-June 2014

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Phylogenetic Tree Construction of Biological Datasets of Cyclooxygenase (COX-1) and (COX-2) by using Cluster Analysis Based on Experimental Values

Chukka Santhaiah
Research scholar, Dept of C.S.E
S.V.U.College of Engg, S.V.University,
Tirupati, A.P, INDIA

Dr.A.Rama Mohan Reddy
Professor, Dept of C.S.E
S.V.U.College of Engg, S.V.University,
Tirupati, A.P, INDIA

Abstract: Phylogenetic trees are worn to symbolize development of associations between biological genus and organisms. The erection of phylogenetic trees is support on the resemblance or dissimilarity of their physical or inherited features. Conventional looms of erecting phylogenetic trees essentially focus on substantial characteristics. The current encroachment of high-throughput knowledge has lead to buildup of enormous quantity of biological data, which in rotate amend the approach of biological studies in a mixture of approaches. This work is mainly focus on constructing the phylogentic tree for Cyclooxygenase of COX-1 and COX-2 based on experimental values. Here to constrct the phylogenetic tree by applying the cluster and by using JavaTree approaches on COX-1 and COX-2. These results are shown the better evolutionary relationship among the COX biological datasets.

Keywords: Phylogenetic tree, Cyclooxygenase, Java Tree, cluster.

I. INTRODUCTION

A phylogenetic tree is a vivid demonstration of the completion connections of genus, and the phylogenetic reserve surrounded by the species replicate the closeness of evolutionary relationships. Conventional erection of phylogenetic trees was essentially based on physical similarities and diversity. Though, the method of the deepness has been changed because of the production of enormous amounts of biological data. For example, high-throughput sequencing expertise have generated genome sequences in numerous thousand organisms. A genomic sequence is fundamentally a thread of four dissimilar kinds of nucleotides (A, C, G and T), with the length from hundreds of thousands to millions. It has been extensively time-honored that the genomic sequences are extremely analogous for evolutionary closed organisms, but not similar for evolutionary distant organisms. So, genomic sequences have been broadly used for building phylogenetic trees [1-3].

The building of phylogenetic trees by means of genomic sequences does have a number of issues. The genomic sequences are frequently long; therefore compare genomic sequences from corner to corner species for building phylogenetic trees is computationally expensive. On the further hand, living organisms in a small position frequently swap over their genetic materials each other, also recognized as straight gene transfer, making it harder to conclude evolutionary relationships based on genomic sequences only. Additionally, present genomic sequence likeness measurement cannot truly reveal evolutionary relationships across the species. Thus, it is necessary to use other data and methods to reveal true relationships [4].

In parallel to the high-throughput genome sequencing technologies, COX data have also been generated in the past decade. The study of using of COX data for biological studies is also known as Cyclooxygenase. The COX data from organisms are very informative since they can reveal internal inflammation mechanisms. Theoretically, evolutionary distant species should have different inflammation activities and patterns, while closely related species should have similar patterns. Therefore, it is desirable to use COX data for phylogenetic exploration, or complement the gene-based phylogenetic exploration to some degree.

COX data have been operated and performed, and corresponding experimental values have been built for scientific communities. On the COX to perform different operations by using cluster analysis. The operations are filter data values it is helpful to eliminate the unwanted genes from the database. Then to operate the values in cluster for adjust the data by applying the log transform data by selecting center genes , normalize genes and center arrays , normalize arrays.

The COX experimental values can be corresponding to as bound for or unbounded graphs. The nodes in graphs can moreover be symbolized as values that are linked by the closed, or be signified as enzymes linked by its values. In consequence, via the in sequence prearranged in the graphs be able to disclose development relationships across the species [5].

II. METHODOLOGY

In this paper, we aim to reveal phylogenetic distances across the species using experimental values, rather than sequence information in the graphs. We use the data of COX experimental values**.** In the relation network, enzymes and genes are represented as nodes, while the substrate and product compounds are represented as edges. The related structural information from the graphs was used for computing phylogenetic distances[6].

A. Description of COX data set:

Cyclooxygenase (COX) is the enzyme that catalyzes the oxidation and subsequent reduction of arachidonic acid to form Prostaglandin G2 and Prostaglandin H2 (PGH2). We collected the processed dataset for COX-1 and COX-2 experimental values of different genes from the NCBI. The

gene names are BG08689, AW555640, BG075861 and BG085367 and these experimental values are changed on different operations like mutations on those genes. These datasets has various values shows in (Table 1) [7].

Table 1: gene names and its experimental values

Gene name	COX-1	COX-2
BG086892	2.40	1.69
AW555640	-5.49	2.16
BG075861	-1.86	1.68
BG085367	-2.84	2.15

B. Data processing:

The flow chart of our approach is outlined in Figure 1, and the detailed description of each step is presented as follows and it has three steps [8].

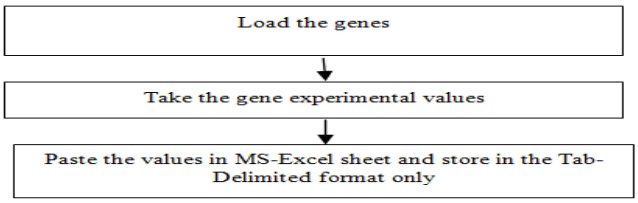

Figure 1: Process of the data

a. ***Load the genes***: For the first step to open the NCBI and enter the gene names i.e its accession numbers.

b. ***To take the Result values:*** The experimental values from the results of COX. The values are indicating the different values in different experiments.

c. ***Store the experimental values:*** The experimental values are stored in excel sheet for further operation of cluster. The values are saving in Tab-Delimited format only because it provides exact results in analysis.

III. CLUSTERING ALGORITHM

Clustering is an unendorsed culture algorithm that finds the concealed arrangement in the unlabeled data. In this work, we used the filter the values, adjust the data values, then apply the hierarchical method, k-means algorithm, self-organizing maps (SOM), and finally apply the Principal Component Analysis (PCA) for avoid the unwanted values, adjust the data with help of log transform, for clustering genes and arrays with hierarchical clustering by centroid linkage, to organize the genes with k-means algorithm, calculate the SOMs and PCA analysis [9]. The results were visualized by Java TreeView [10].

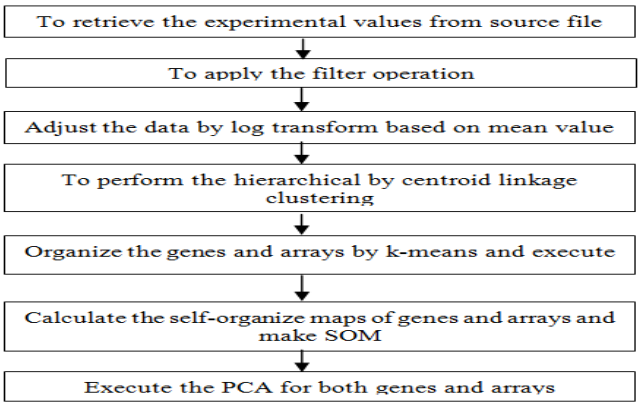

Figure 2: The execution of clustering process

IV. RESULTS

The above clustering is performed different operations on COX gene values. In the methodology I select the four genes and its values for clustering operations. The method is provides good results for analyze of phylogenetic tree by using the JavaTree. The output of cluster analysis especially cox.cdt is helpful in JavaTree for constructing the phylogenetic tree.

We have applied Cluster 3.0 on biological value datasets, and used Java TreeView to generate the dendrograms (or phylogenetic trees) for each of the dataset. Figure 3 show the dendrograms of COX genes with the lengths of the branches reflecting the distances between species. Therefore, the shorter the branches, the evolutionarily closer the species are, and the longer the branches, the evolutionarily more distant the species are.

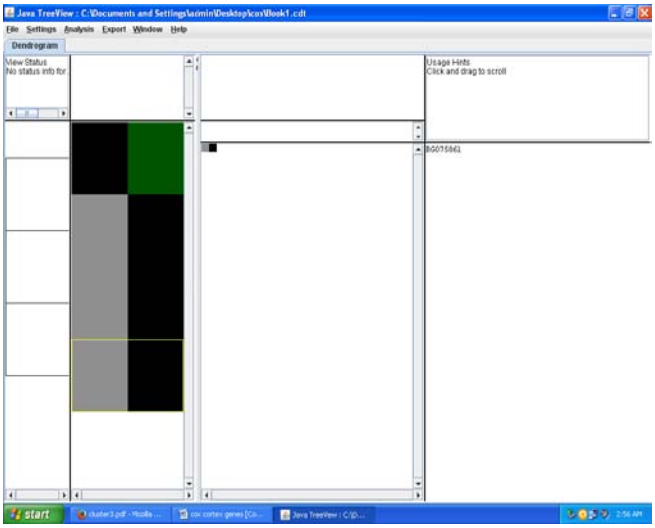

Figure 3: The dendogram of gene BG075861

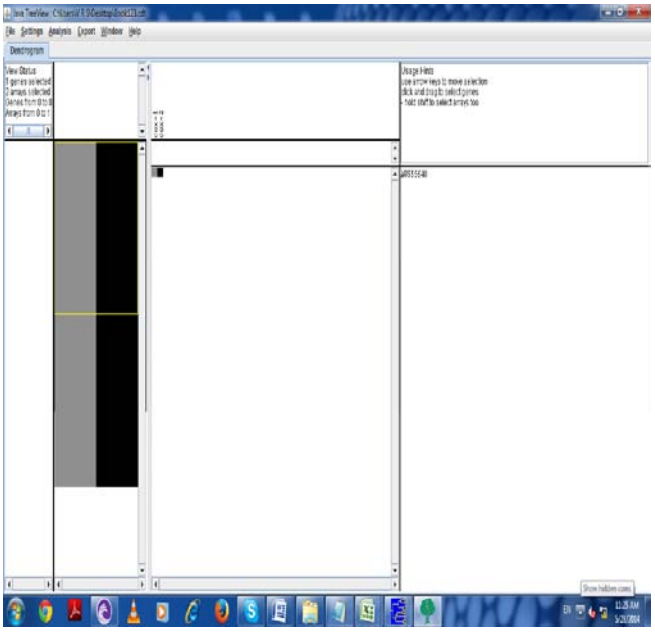

Figure 4: Dendogram gene of AW55640

From the phylogenetic tree in (Figure 3), we preserve discover a quantity of fascinating consequences. For illustration, we can see the contiguous species by means of human beings is *Mus musculus*, supposed domicile mouse. To appreciate why these two species stay close, we did

literature search about the closeness of these two species. We found that almost all genes in the mouse were also present in humans. Actually, researchers have reported that approximately 99% of mouse genes have counterparts in humans . The dendrogram result for the COX also strongly indicates the effectiveness of our approach.

V. CONCLUSION

In this dissertation, we have statement our loom that uses the in sequence of COX reactions to reveal the next of kin of progress between species. The foremost contribution of this study is to demonstrate that with the usage of clustering, the phylogeny of species can be constructed by a higher level function. Our experimental results have shown that our approach is pretty accurate in most of cases, strongly indicating that effectiveness of our approach.

VI. #REFERENCES

[1] Snel B *et al. Nat Genet*. 1999 21: 108 [PMID: 9916801]

[2] Li Ming *et al. Bioinformatics.* 2001 17: 149 [PMID: 11238070]

[3] Herniou EA *et al. Journal of virology*. 2001 75: 8117 [PMID: 11483757]

[4] Ma HW & Zeng AP, *Bioinformatics.* 2003 19: 270 [PMID: 12538249]

[5] Smith WL, DeWitt DL, Garavito RM. Cyclooxygenases: structural, cellular, and molecular biology. Annu Rev Biochem. 2000; 69:145–182. doi: 10.1146/annurev.biochem.69.1.145.

[6] Kam PC, See AU. Cyclo-oxygenase isoenzymes: physiological and pharmacological role. Anaesthesia.2000; 55:442–449. doi: 10.1046/j.1365-4.2000.01271.x. [PubMed] [Cross Ref]

[7] http://genomebiology.com/2007/8/1/R14/table/T1.

[8] Jain, A. K., and Dubes, R. C. (1988). *Algorithms for* clustering data (Englewood Cliffs,N.J.: Prentice Hall)

[9] de Hoon MJ *et al*. *Bioinformatics*. 2004 **20**: 1453 [PMID: 14871861]

[10] Saldanha AJ, *Bioinformatics*. 2004 20: 3246 [PMID: 15180930]