International Journal of Engineering Research and Development
e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 10, Issue 6 (June 2014), PP.26-32

Constructing of Phylogenetic Tree for COX based on sequences by using ClustalW

Chukka Santhaiah[1], Dr.A.Rama Mohan Reddy[2]
[1]*Research Scholar,C.S.E Department, S.V.University , Tirupati , A.P, INDIA.*
[2]*Professor, C.S.E Department, S.V.University , Tirupati , A.P, INDIA.*

Abstract:-The phylogenetics development of genus is regularly described as the consequence of random mutations and expected choice, which gives rise to the perception of phylogenetic trees. The big quantity of soaring excellence sequence data accessible has ended the rebuilding of phylogenetic trees as of sequence data achievable. In this work COX gene is taken from NCBI P22437.1 as a protein with its accession number. The Clustal approach is a influential and admired alternative for create phylogenetic trees as of sequence data, no limit in the size of the trees, which can be constructed. The aim of work is speed up tree reconstruction by construction use of preceding information. Here put forward a new heuristic for incorporating a phylogenetic tree for protein datasets are converted to DNA datasets and construct the phylogenetic tree for analogous data set. In this work we implement the proposed Clustal for multiple alignments. The experiments performed show, which incorporating prior knowledge leads to a significant speed up, if large amounts are included.

Keywords:- Clustal, DNA, Protein, Phylogenetic tree

I. INTRODUCTION

Phylogenetics is the cram of development affairs linking organisms. Expertise has permit astonishing evolution in phylogenetics. One portion of this in the biology itself. The innovation of DNA and the capability for biologists to sequence DNA has, to say the slightest revolutionized the field. Computers have assisted extremely as well. To the side from their relieve in sequencing assignments, computers smooth the progress of a broad array of phylogenetic problem-solving activities.

The contribution of computers (or perhaps more appropriately, computer science) is text processing. Since DNA sequences come as a sequence of characters (from the alphabet A, C, T, G), there are several computer science algorithms that can be used as processing tools. The field of phylogenetics has applications to molecular biology, genetics, evolution, epidemiology, ecology, conservation biology, and forensics to name a few. Phylogenies are the chronological and evolutionary relationships among organisms. Researchers can employ this data to better understand how viruses spread or to study common biological processes between different species of life.

Cyclooxygenase also known as Prostaglandin endoperoxide H synthase (PGHS, EC.1.14.99.1) and exists in two isoforms; (COX-1) and (COX-2), which catalyses the oxidation of AA to prostanoids. COX-1 and COX-2 enzymes are heme proteins, homodimers that are widely distributed. These enzymes are located in the lumenal portion of the endoplasmic reticulum membrane and the nuclear envelope [1]. Cox isoenzymes are also involved in a wide range of pathologies that include for Cox-1 thrombosis, and for Cox-2 inflammation, pain and fever, various cancers, and neurological disorders like Alzheimer's and Parkinson's diseases [2]. Big success was achieved with the development of non-steroidal anti-inflammatory drugs (NSAIDs) such as aspirin for treatment of fever and pain [3].

A multiple sequence alignment (MSA) is a sequence alignment of three or more biological sequences such as protein, DNA, or RNA. Typically it is implied that the set of equences share an evolutionary relationship, which means they are all descendents from a common ancestor. These regions may correspond to functional, structural, or evolutionary relationships between the sequences. Alignments can reflect a degree of evolutionary change between sequences that are descendants from a common ancestor. There is a relationship between phylogenies and sequence alignments [4].

There are various techniques used to create phylogenetic trees and most of them rely on aligned genetic sequences to perform this task. Probably the most popular genetic sequence alignment algorithm is ClustalW [4]. Although successful in its domain, ClustalW is very sensitive to highly divergent sequences. Therefore the purpose of this project is to modify the ClustalW sequence alignment algorithm so that it can be used to construct a more accurate tree when highly divergent sequences are present. Sequence alignment is a way of arranging sequences of DNA, RNA, or proteins in order to distinguish regions of similarity. ClustalW is a popular program used for multiple sequence alignment and for preparing phylogenetic trees. Its portability amongst various computing platforms is the main reason for its widespread use. Due to its popularity and the

availability of source code, ClustalW was used for this project. The progressive alignment algorithm used by ClustalW to perform a multiple sequence alignment [5].

II. METHODOLOGY

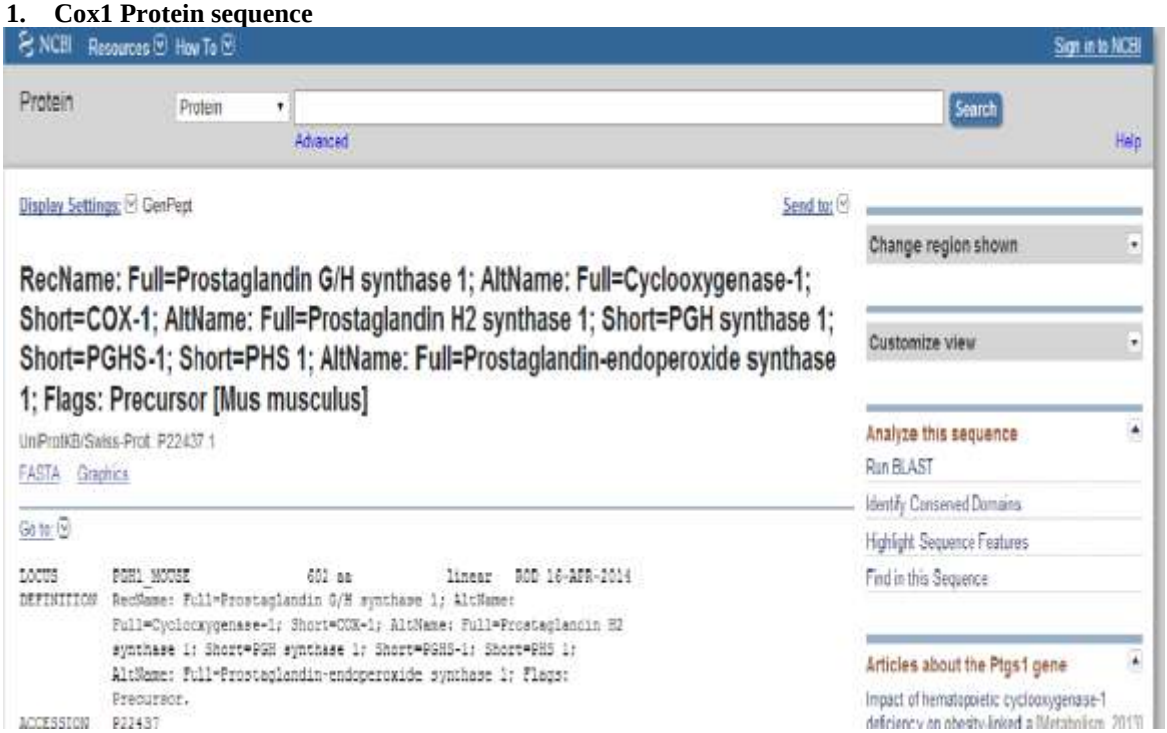
The methodology purpose is to get the nucleotide sequences by using protein sequences that are related to same sequences. In this process first select the one protein that is main protein for entire this process. Here that protein P22437.1. It is a protein that is related to COX gene of the one family. To get the nucleotide sequence from amino acid sequence by performing different operations on selected protein. The methodology follows different operations.

a. Search the protein form database that is related to COX.
b. Give that protein in the NCBI database for find the sequence in the form of FASTA [6].
c. Then select the FASTA format of protein and paste the in BLAST for BLAST operation of tBLASTn. It takes some time for execution based on sequence length [7].
d. The output is gives graphic summary and description of gene. And it displays related genes.
e. In output contains the gene description of the related name of genes, maximum score, total score identity of genes and accession number of that genes.
f. Based on gene accession number to find the nucleotide sequences from protein sequences.
g. Then select the all sequences, copy the FASTA format of the sequences and paste in ms-word format for ClustalW operation.
h. Finally perform the ClustalW operation and it provides rooted and Unrooted phylogenetic tree [8].

In the above operations we get the nucleotide sequence from amino acid sequence, and also to create the phyologenetic tree based on nucleotide sequences. To get the phylogenetic tree with the help of DNA sequences. The output file contain rooted and unrooted phylogenetic tree. In phylogenetic tree the similar sequences are form as taxa. The remaining sequences are grouped as a whole tree of similar tree. The execution of the nucleotide process may take few seconds/minutes based on sequence length, processor and internet speed.

III. RESULTS

1. Cox1 Protein sequence



>gi|129900|sp|P22437.1|PGH1_MOUSE RecName: Full=Prostaglandin G/H synthase 1; AltName: Full=Cyclooxygenase-1; Short=COX-1; AltName: Full=Prostaglandin H2 synthase 1; Short=PGH synthase 1; Short=PGHS-1; Short=PHS 1; AltName: Full=Prostaglandin-endoperoxide synthase 1; Flags: Precursor
MSRRSLSLWFPLLLLLLLPPTPSVLLADPGVPSPVNPCCYYPCQNQGVCVRFGLDNYQCDCTRTGYSGP
NCTIPEIWTWLRNSLRPSPSFTHFLLTHGYWLWEFVNATFIREVLMRLVLTVRSNLIPSPPTYNSAHDYIS
WESFSNVSYYTRILPSVPKDCPTPMGTKGKKQLPDVQLLAQQLLLRREFIPAPQGTNILFAFFAQHFTHQ

FFKTSGKMGPGFTKALGHGVDLGHIYGDNLERQYHLRLFKDGKLKYQVLDGEVYPPSVEQASVLMRY
PPGVPPERQMAVGQEVFGLLPGLMLFSTIWLREHNRVCDLLKEEHPTWDDEQLFQTTRLILIGETIKIVI
EEYVQHLSGYFLQLKFDPELLFRAQFQYRNRIAMEFNHLYHWHPLMPNSFQVGSQEYSYEQFLFNTSM
LVDYGVEALVDAFSRQRAGRIGGGRNFDYHVLHVAVDVIKESREMRLQPFNEYRKRFGLKPYTSFQEL
TGEKEMAAELEELYGDIDALEFYPGLLLEKCQPNSIFGESMIEMGAPFSLKGLLGNPICSPEYWKPSTFG
GDVGFNLVNTASLKKLVCLNTKTCPYVSFRVPDYPGDDGSVLVRRSTEL

Fig 1: The FASTA format of input Protein (P22437.1)

2. Cox1 Protein sequence tBLASTn results

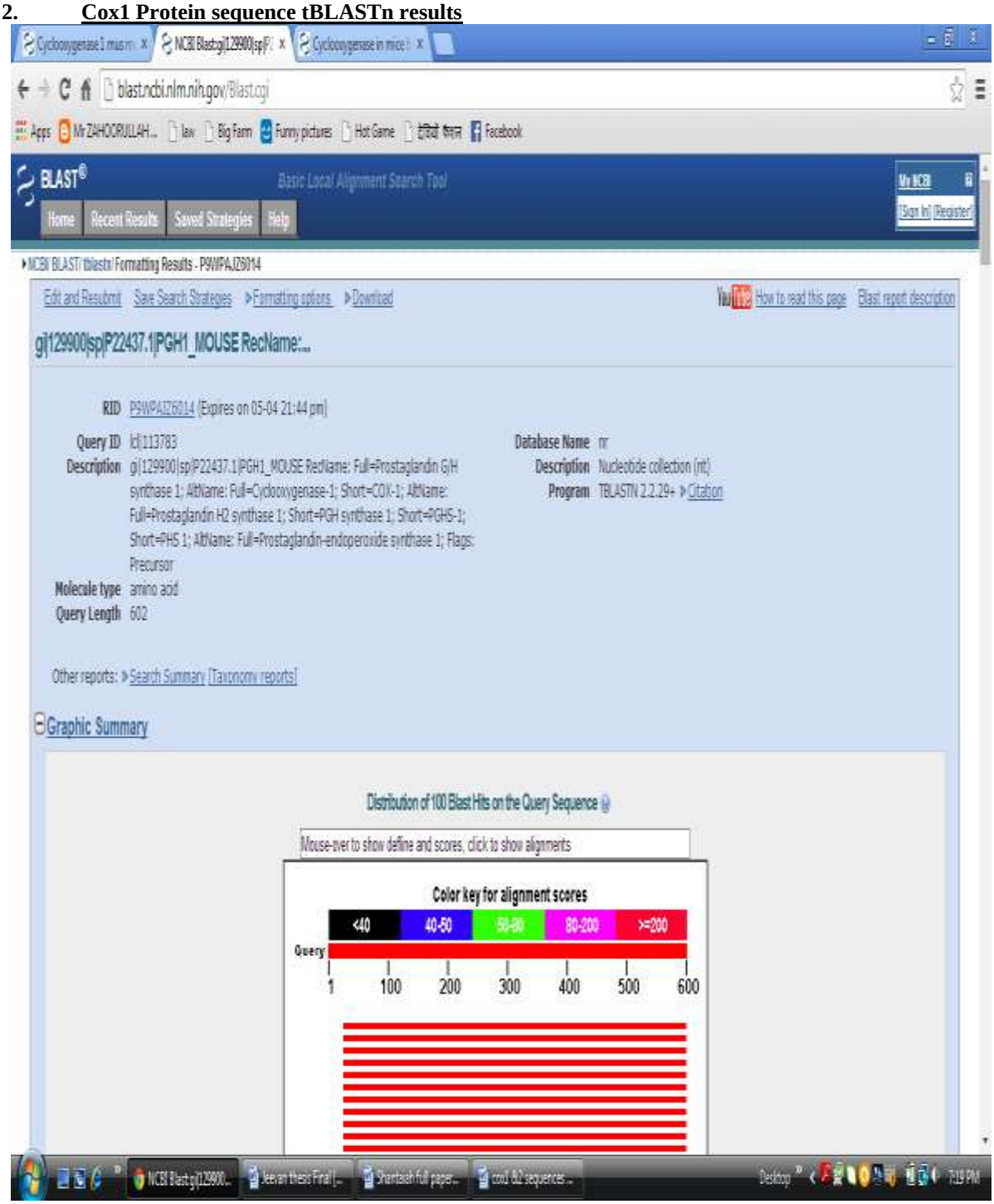


Fig 2: The graphic Summary of query sequence.

ClustalW Results:

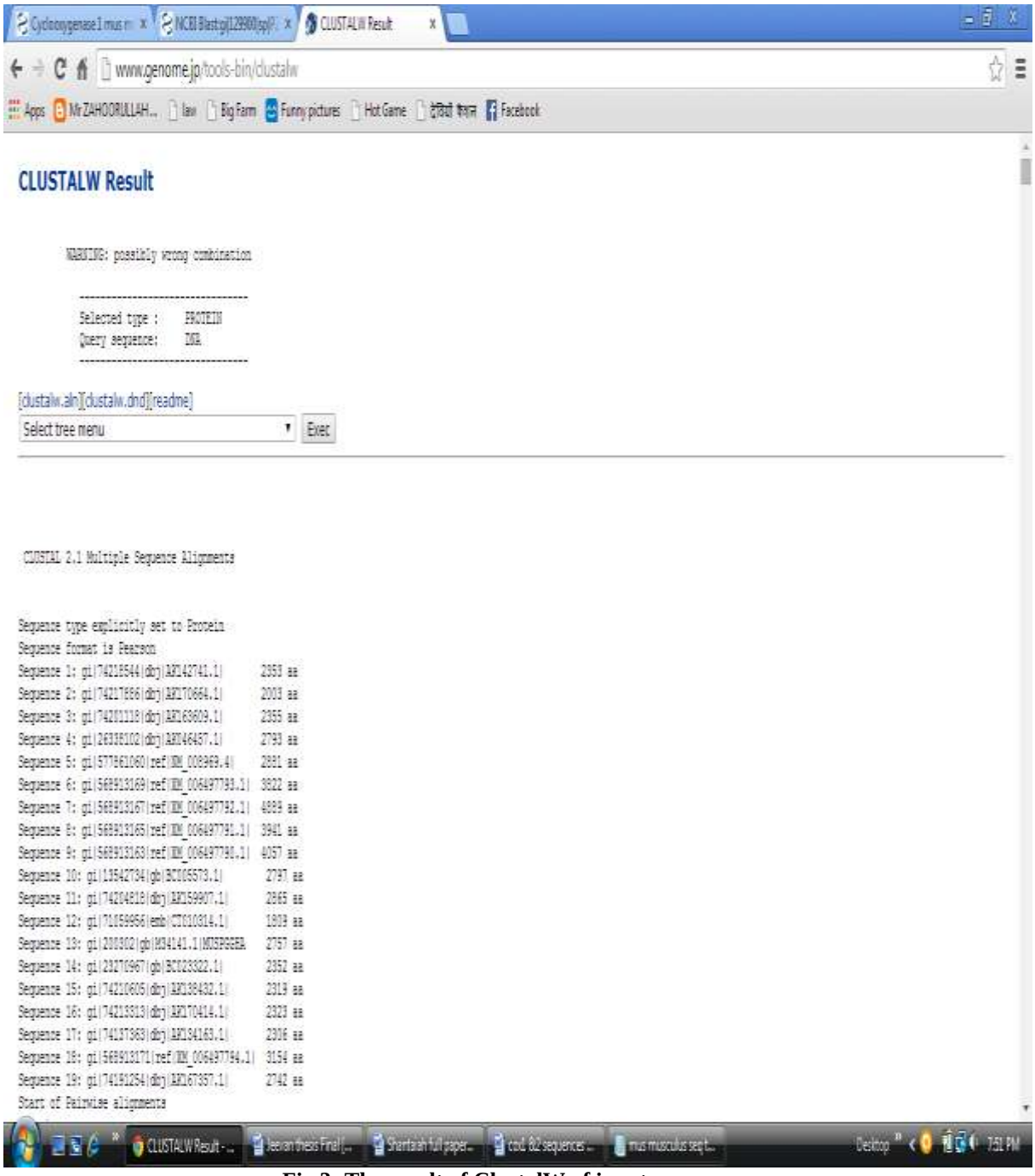


Fig 3: The result of ClustalW of input sequences

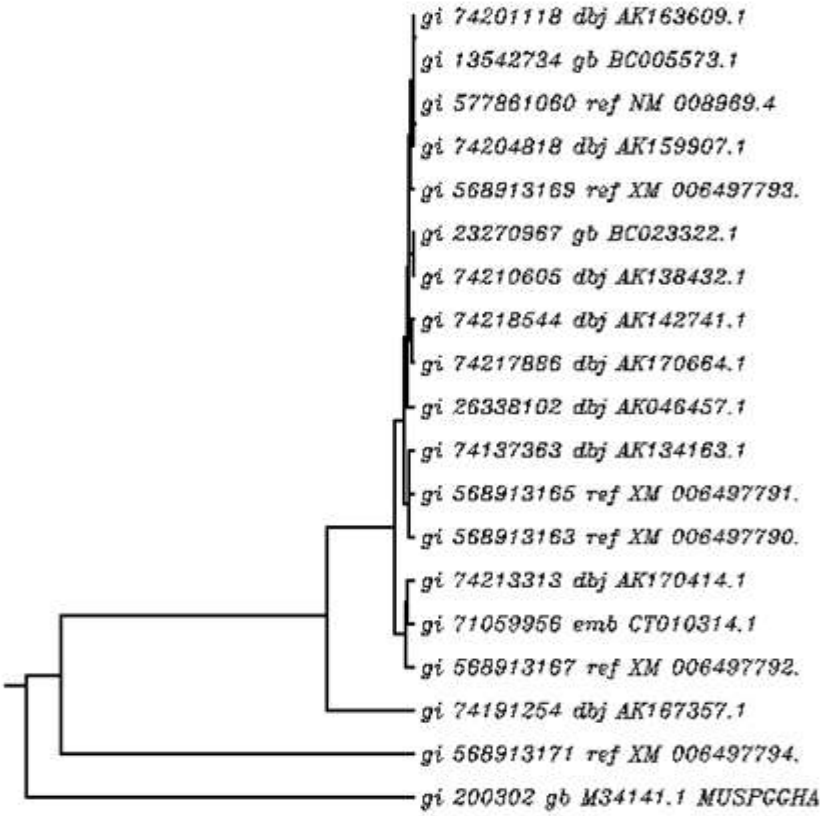


Fig 4: ClustalW result: Rooted Phylogenetic Tree with branch length (UPGMA) of Cox1

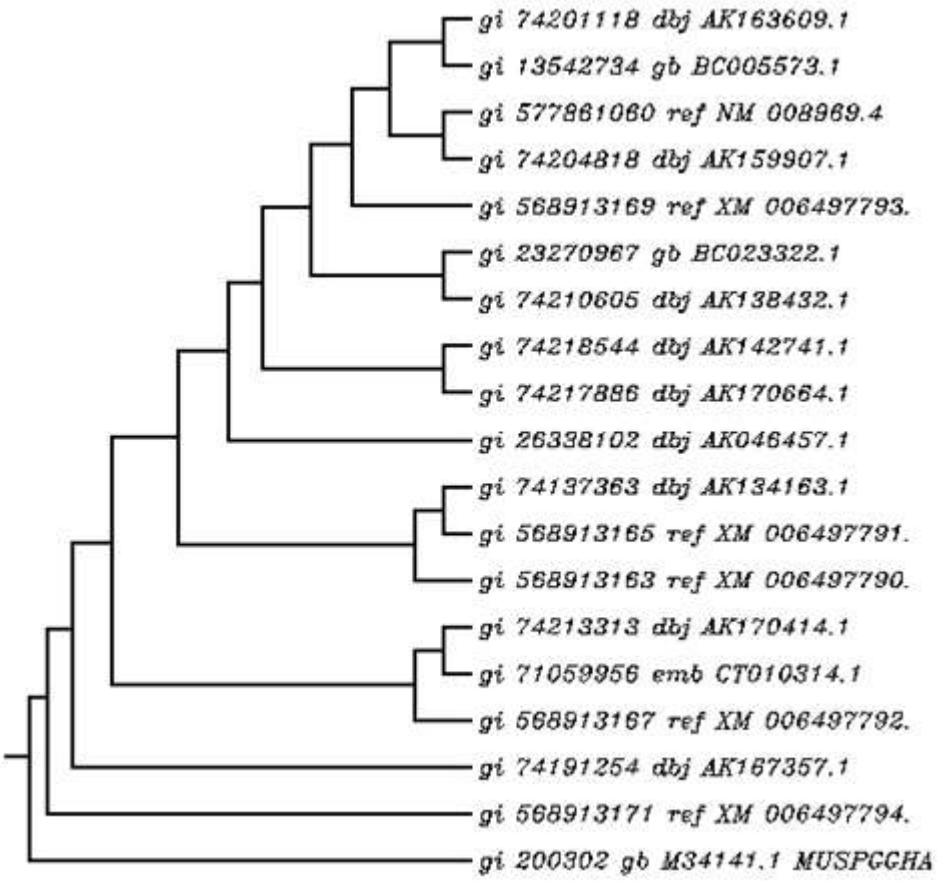


Fig 5: ClustalW result: Rooted Phylogenetic Tree (UPGMA) of Cox1

Cox2 Protein sequence

>gi|13487701|gb|AAK27680.1| cyclooxygenase 2 [Mus musculus]
XTRQIAGR

Cox2 Protein sequence tBLASTn results

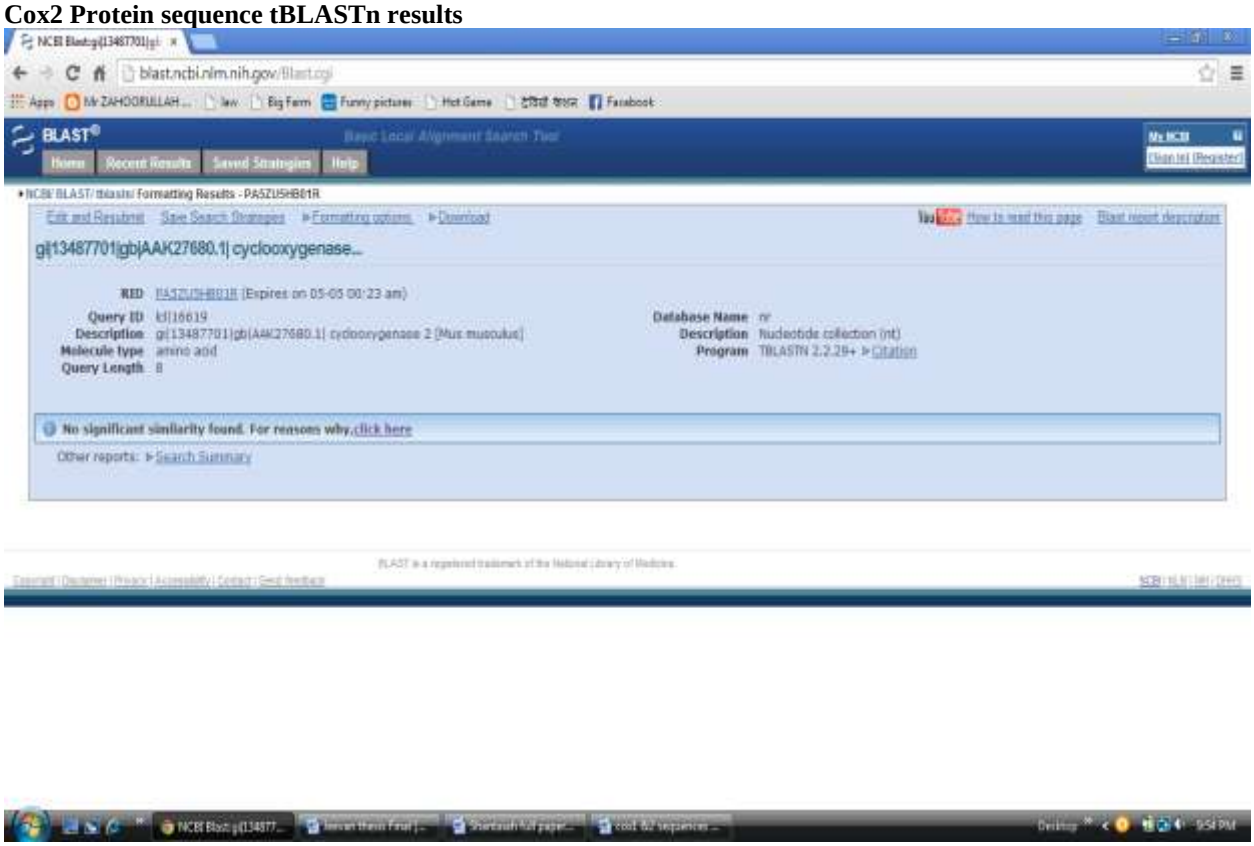


Fig 6: No significant similarity of input sequence (AAK27680.1)

IV. CONCLUSIONS

It has been shown to a great extent curiosity in phylogenetic systematic recently. Since the reason that it is a technique for biologists to rebuild the pattern of events that have led to the distribution and diversity of life. Due to the "Tree of Life" proposals, and studies the researchers have tried to find some new methods to combine large number of trees to construct phylogenies on hundreds, or even thousands of species. For understanding the language of genes and proteins we have to find a suitable model of how they have evolved in the course of evolution. Because of this we need to develop tree building methods which discover the process of evolution. These kinds of methods have gained importance with the advent of molecular biology.

REFERENCES

[1]. J. R. Vane,Y. S. Bakhle1, and R. M. Botting , Annu. Rev. Pharmacol. Toxicol. 1998. 38:97–120, Cyclooxygenases 1 and 2

[2]. Garavito RM, Malkowski MG, DeWitt DL. 2002. The structures of prostaglandin endoperoxide H synthases-1 and -2. Prostaglandins Other Lipid Mediat 68-69:129-52.

[3]. Mag.rer.nat.Kerstin Kitz ,2011,Transcriptional Regulation of Cox-2 Expression in Human Osteosarcoma cells.

[4]. Higgins,D.G. (1994) CLUSTAL V: multiple alignment of DNA and protein sequences. Methods Mol. Biol., 25, 307–318

[5]. Thompson, Julie D., Desmond G. Higgins, and Toby J. Gibson. "CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment Through Sequence Weighting, Position Specific Gap Penalties and Weight Matrix Choice." Nucleic Acids Research 22 (1994): 4673-4680.

[6]. http://www.ncbi.nlm.nih.gov/

[7]. http://www.ncbi.nlm.nih.gov/tblastn

[8]. http://www.genome.jp/clustalw/