Molecular Diversity And Interrelationship Amongst The Various Phyla Based On 16s And 12 Srrna Analysis

Debojyoti Dutta

Abstract: Sequence alignment is the fundamental and best approach for the phylogenetic study and to infer the sequence similarity and dissimilarity between various species. Sequence alignment can be done with the amino acids or nucleic acids, but due to redundancy of genetic code some information always hidden in case of amino acid sequencing, but with nucleic acid sequencing this problem is not appear. Nucleic acid sequence alignment can indicate the molecular diversity and also can tell us about the conserved sequences throughout the different organisms from phylum Porifera to Vertebrata. For these purpose different organisms mitochondrial 12SrRNA is collected from NCBI data base and with the help of different bioinformatics tools all the sequences are analyzed. This paper attempt to establish the significance of mitochondrial12SrRNAs in phylogenetic studies, and a phylogenetic tree is also constructed with the help of Multiple Sequence Alignment (MSA) tools. Some highly conserved sequences were identified and these sequences can be used to form a primer with more advanced techniques, but in this paper only sequence collected, analyzed and relationship between them established. By extensive data mining, sequence similarity searches of databases give all the information about the query and similar sequences. Information about these sequences helps to predict the relation between the sequences as well as organisms and form the basic knowledge about the molecular evolution though out different phyla.

keywords: Sequence Alignment, Mitochondrial 12S rRNA, NCBI Data Base, Phylogenetic Tree

1. INTRODUCTION:-

In Biological science, diversity is the term which can indicate the differences between genes and as consequences of these differences in genes variability in species occurs. The evolutionary history of thousands of species uncovers a fundamental truth: genetic variation plays the major role in speciation as well as in evolution. So these variations can help us to build a molecular basis of speciation based on the nucleic acid analysis. These analysis results also indicate the sequence similarity and dissimilarity between organisms and help us to build a phylogenetic tree. Protein based phylogenetic study includes the amino acid sequencing, gel electrophoresis (including IEF), protein interactions (antigenic distance) etc. but these all possess some major problems like cross reaction and the techniques are also time consuming results are not also specific [4]. Detection of distantly related sequences is easier in amino acid sequencing, because the redundancy of the genetic code of 64 codons is reduced to 20 distinct amino acids, the functional building block of the proteins. However the loss of degeneracy at this level is accompanied by a loss of information that relates more directly to the evolutionary process, because proteins are a functional abstraction of genetic events that occur in DNA. But the nucleic acid based analyses are specific, sensitive and reliable [2]. Now days phylogenetic analysis is based on multiple DNA-RNA based techniques. DNA based approaches includes DNA hybridization, Restriction Fragment Length Polymorphism: RFLP, Polymerase Chain Reaction: PCR based techniques, Species specific PCR primer use, and DNA sequencing, RNA sequencing.

- *Debojyoti Dutta Department of Zoology A.B.N Seal College*
- *P.O &Dt-Cooch Behar, PIN-736101, West Bengal*
- *(Corresponding author Email-debojyotidutta2001@gmail.com)*

All these techniques are highly dedicated to their discriminatory powers and with high output results. One of the great achievements of modern biology has been the development of accurate and reliable technologies for the rapid screening of DNA sequence variations [3]. With help of these in-silio studies this paper demonstrate the importance of the mitochondrial 12S r-RNA sequencing, combined with bioinformatics, for the detection and identification of molecular relationship between the various species [1]. Mitochondrial 12S r-RNA has been used extensively due to high copy number of mitochondria in cell. Mitochondria follow clonal inheritance as only mother to contribute to mitochondria; its genome does not undergo recombination; thus, genetic material will be passed onto the next generation unchanged. Reports also suggest that mitochondrial genome is accumulating high percentage of neutral mutations which is helpful in species identification. With few exceptions, the cells of all eukaryotic species contain mitochondria [5,6] . The mitochondrial DNA is double stranded approximately 16kb in length and which is 1%-2% of total DNA in mammalian cells. Mitochondrial DNA encodes several important protein compounds for the respiratory chain complex: seven subunits of Complex I, one subunit of Complex III (Cytb), three subunits of Complex IV, and two subunits of Complex V. It also encodes two rRNAs (12S rRNA and 16S rRNA) and tRNAs that are required for mitochondrial protein synthesis. Mitochondria possess their own organelle-specific DNA replication, transcription, and translation systems. Due to the above reasons this paper try to solve the molecular relationship between various species with the help of mitochondrial 12S rRNA. Total 18 different organisms are used for this study from different phylum or order, and with the help of proper databases such as NCBI the mitochondrial rRNAs are downloaded as per need and with computational approach the sequences are analysed, and a phylogenetic tree is constructed with the help of NCBI BLAST.

2. METHODOLOGY:-

2.1. Experimental Design:-

To know the sequence divergence and for the contraction of a phylogenetic tree, first of all we need to collect the respective sequences. For this study mitochondrial 12S rRNA is selected. Then from the NCBI data base different 12S rRNAs are selected from different organisms which belong from different orders or phyla. After all the sequences are collected a table is formed to sum up all information. After that for detection of sequence divergence or similarity all the sequences from vertebrate phyla and invertebrate phyla were at first BLAST separately and construct a phylogenetic tree. Then a combined BLAST was perform to show the total sequence divergence across the phyla, and also build a phylogenetic tree with the help of NCBI BLAST and CLUSTRAL OMEGA. From these results and with extensive data base searching attempts are undertaken to show the evolutionary conserved sequences across the different phyla or orders, from which universal primer designing for all these sequences may be possible.

2.2. Sequence Analysis protocol:-

First of all we selected different organisms from different phyla, more accurately from different orders. Then as the 12S rRNA is our choice of interest we collect the respective organisms 12S rRNA from the NCBI DATA BASE. For this collection at first we need to enter the NCBI DATABASE (https://www.ncbi.nlm.nih.gov/). After selecting nuceltide of our interest i.e., 12SrRNA we need to write the specific sequence of demand such 12S rRNA followed by organism name (scientific name) in the main search bar. By clicking the sequences as per need in most cases for sequence alignments FASTA format is collected. In this way total 18 different sequences are collected from different organism (table 1) and it contains the valuable information regarding the sequences.

Table 1:-*Sequences under investigation with proper reference*

Serial No.	Name of the organism	PHYLUM	CLASS	Type of nucleotide sequence	Version	Sequence
1	Oscarella microlobata	Porifera	Calcerea	12S rRNA	NC014850.1	AAA…CAT
2	Heteractis crispa	Cnidarian	Anthozoa	12S rRNA	KC812140.1	TAG…TAC
3	Beroe forskalii	Ctenophore	Nuda	12S rRNA	MG655623.1	CAG…GGT
4	Caenorhabditis elegans	Nematoda	Chromadorea	12S rRNA	JF896455.1	TAA...ATA
5	Megascolex sp.	Annelida	Clitellata	12S rRNA	JX067991.1	AAC…ATG
6	Heterololigo bleekeri	Mollusca	Cephalopoda	12S rRNA	NC_002507.1	TTA…ATA
7	Arbacia lixula	Echinodermata	Eleutherozoa	12S rRNA	X80396.1	GCC…AAT
8	Drosophila melanogaster	Arthropoda	Insecta	12S rRNA	KJ947872.2	TTA…TGA
9	Homo sapiens	Vertebrata	Mammalia	12S rRNA	NR_137294.1	AAT…AAC
10	Mus musculus	Vertebrata	Mammalia	12S rRNA	V00711.1	AAA…AAT
11	Danio rerio	Vertebrata	Actinopterygii	12S rRNA	AC024175.3	CAA…ATC
12	Xenopus laevis	Vertebrata	Amphibia	12S rRNA	M10217.1	TAA…AAT
13	Naja naja naja	Vertebrata	Lepidosauria	12S rRNA	GQ225677.1	CCA…AAT
14	Gallus gallus	Vertebrata	Aves	12S rRNA	AB086102.1	AAA…TAC
15	Duttaphrynus melanostictus	Vertebrata	Amphibia	12S rRNA	KJ001636.1	CAA…AGG
16	Rhinoceros unicornis	Vertebrata	Mammalia	12S rRNA	NC_001779.1	CAT…AAT
17	Hoolock leuconedys	Vertebrata	Mammalia	12S rRNA	NC_033882.1	AAT…AAC
18	Trachypithecus pileatus	Vertebrata	Mammalia	12S rRNA	NC_024529.1	ATA…AAT

3. BLAST RESULTS:-

3.1. For the 8 invertebrate representative organisms:-

At first 8 different invertebrate organisms sequences are blast with NCBI BLAST and CLUSTRAL OMEGA for multiple sequence alignment purpose. In case of NCBI BLAST, at first we take the Oscarella microlobataas our query sequence (Table 2) and other 7 sequences (table 3) is treated as subject sequences.

Table 2:- *Query Sequence information*

Query ID	Organism Name	Base Pair (bp)
lcl|Query_1283	Oscarella microlobata	1283

Table 3:- *List of multiple subject sequences information*

Query ID	Organisms name	Base pair
lcl|Query_2145	Caenorhabditis elegans	697
lcl|Query_2146	Drosophila melanogaster	786
lcl|Query_2147	Arbaxia lixula	886
lcl|Query_2148	Megascolex sp.	385
lcl|Query_2149	Loligo bleekeri	782
lcl|Query_2150	Heteractis crispa	788

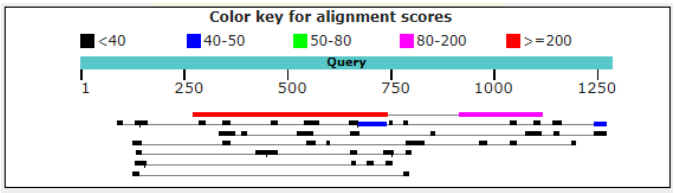


Fig. 1 *:- Distribution of the top 59 BLAST hits on 7 subject sequences*

The graphic (fig.1) is an overview of the database sequences aligned to the query sequence. These are represented horizontal bars colour coded by score and

showing the extent of the alignment on the query sequence. Separate aligned regions on the same database sequence are connected by a thin grey line. Subject sequences and query sequences show poor similarity (table 4).

Table 4:- *Sequences Producing Significant Alignments*

Organisms	*max score*	*total score*	*query cover*	*e-valu*	*percent identity*	*accession*
Heteractis crispa	*343*	*450*	*52%*	*6e-97*	*77.00 %*	*Select seq lcl|Query_21 50*
Drosophila melanogaster	*44.6*	*512*	*26%*	*2e-07*	*88.24 %*	*Select seq lcl|Query_21 46*
Arbaxia lixula	*37.4*	*226*	*16%*	*3e-05*	*80.00 %*	*Select seq lcl|Query_21 47*
Loligo bleekeri	*31.0*	*214*	*13%*	*0.00 5*	*75.00 %*	*Select seq lcl|Query_21 49*
Beroe forskalii	*25.6*	*178*	*9%*	*0.21*	*93.75 %*	*Select seq lcl|Query_21 51*
Caenorhabditis elegans	*22.9*	*180*	*5%*	*0.73*	*81.82 %*	*Select seq lcl|Query_21 45*
Megascolex sp	*21.1*	*42.2*	*2%*	*2.5*	*87.50 %*	*Select seq lcl|Query_21 48*

No highly similar sequence is detected between these organisms, although some similarities also found, on the basis of these somewhat similar sequences a phylogenetic tree is constructed by NCBI BLAST and also for CLUSTRAL OMEGA.

PHYLOGENETIC TREE BY NCBI BLAST:-

This tree is form with NCBI BLAST by Neighbor Joining with distance correction method (Fig. 2) . But with the help of CLUSTRAL OMEGA (Fig.3) without distance correction a tree is also generated, which includes all the organisms regardless of too much dissimilarity.

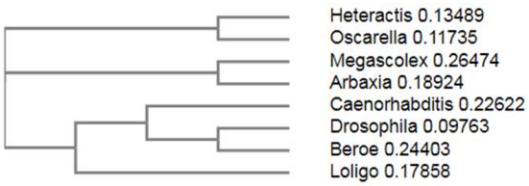


Fig.2*. Neighbour-joining tree without distance corrections among invertebrate species.*

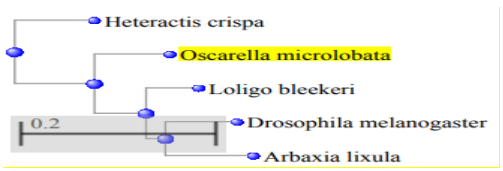


Fig.3*. Phylogenetic tree among studied invertebrate species with the help of Clustral omega.*

Multiple Sequence Alignment Viewer 1.10.0. Total 874 base pair displayed. No particular similar sequences found within the invertebrate series. Discrete pattern of similarity within some regions found. Colour red indicating the regions of dissimilarity. Colour grey regions of similarity. Very few regions of similarity within these sequences are in grey colour.

3.2. MEGABLAST FOR INVERTEBRATE SERIES:-

In case of invertebrate series we take Oscarella microlobata's 12S mitochondrial rRNA as our query sequence and other selected 7 organisms 12S mitochondrial rRNA as subject sequences (Table 5). The MEGABLAST is performed for identification of highly similar sequences between these organisms (Fig 4).

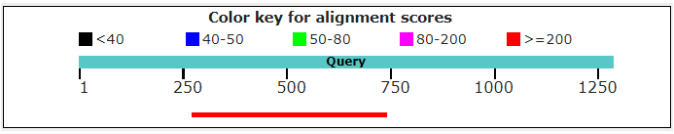


Fig.4 :- *Distribution Of The Top 1 Blast Hits On 1 Subject Sequences*

Only one out of the 7 subject sequences shows highly similar sequence with the query sequence. Others organisms simply shows no highly similar sequences. It shows 77.97% sequences similarity with the query sequence, within the 36% query cover areas.

Table 5 :- *Result of Megablast among studied invertebrates*

Organism	*Max. score*	*Total score*	*Query Cover*	*E-Value*	*Percent Identity*	*Accession*
Heteractis crispa	*279*	*279*	*36%*	*3e-78*	*77.97%*	*Query_53230*

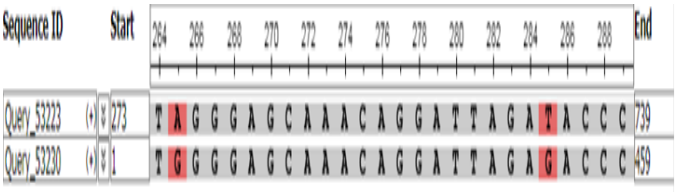


Fig.5 *Multiple Sequence Alignment Viewer 1.10.0.*

This graphic (Fig.5) shows the subject sequences which has the highly similar sequences with the query sequences. Only one subject sequence (Heteractis crispa) shows highly similar sequence with the query sequence. This result of megablast on the invertebrate organisms shows no overall highly similar sequences. The MSA viewer shows the both sequences with red mark indicating the difference in base pair.

3.3. PHYLOGENETIC TREE ON THE BASIS OF MEGABLAST:-

So from the above results of megablast it is clear those only 2 sequences shows highly similar sequences with each other. Others are not showing sequences similarity at this level and also excluded from the phylogenetic tree (Fig. 6).

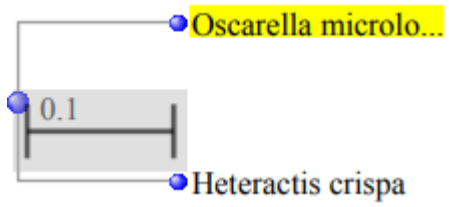


Fig.6*. Phylogenetic tree based on MegaBlast.*

(a) For the 10 vertebrate organisms:-

In case of vertebrate series the Homo sapiens 12S mitochondrial ribosomal RNA is act as a query sequence (Table 6).

Table 6 :- *Query sequence information*

Query Id	*Organism Name*	*Base pair*
lcl|Query_114117	*Homo sapiens*	*954*

Other 9 sequences are now treated as a subject sequence (Table 7).

Table 7 - *List of subject sequences*

Query id	*Organism Name*	*Base Pair*	*Query id*	*Organism Name*	*Base Pair*
lcl|Query_144119	*Mus musculus*	*955*	*lcl|Query_144124*	*Duttaphrynus melanostictus*	*668*
lcl|Query_144120	*Danio rerio*	*952*	*lcl|Query_144125*	*Rhinoceros unicornis*	*971*
lcl|Query_144121	*Xenopus laevis*	*819*	*lcl|Query_144126*	*Hoolock leuconedys*	*953*
lcl|Query_144122	*Naja naja naja*	*930*	*lcl|Query_144127*	*Trachypithecus pileatus*	*947*
lcl|Query_144123	*Gallus gallus*	*976*			

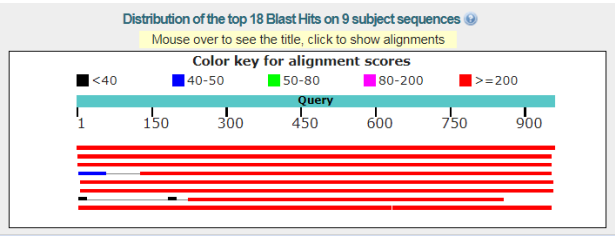


Fig.7. *Overall similarities and dissimilarities among studied vertebrate species.*

The graphic (Fig.7) is an overview of the database sequences aligned to the query sequence. These are represented horizontal bars colour coded by score and showing the extent of the alignment on the query sequence. Separate aligned regions on the same database sequence are connected by a thin grey line. The query and subject sequences show highly similar sequences (Table 8).

Table 8:- *Sequences Producing Significant Alignments*

Organisms	*Max score*	*Total score*	*Query cover*	*e-valu*	*Percent identity*	*Accession*
Hoolock leuconedys	*1265*	*1288*	*100 %*	*0.0*	*89.64 %*	*Select seq lcl|Query_114126*
Trachypithecus pileatus	*1003*	*1068*	*99%*	*0.0*	*84.38 %*	*Select seq lcl|Query_114127*
Rhinoceros unicornis	*755*	*755*	*99%*	*0.0*	*78.17 %*	*Select seq lcl|Query_114125*
Mus musculus	*628*	*672*	*91%*	*0.0*	*77.74 %*	*Select seq lcl|Query_114119*
Danio rerio	*398*	*420*	*98%*	*1e-113*	*70.38 %*	*Select seq lcl|Query_114120*
Gallus gallus	*394*	*394*	*98%*	*1e-112*	*70.86 %*	*Select seq lcl|Query_114123*
Xenopus laevis	*281*	*339*	*69%*	*2e-78*	*71.32 %*	*Select seq lcl|Query_114121*
Duttaphrynus melanostictus	*215*	*237*	*65%*	*8e-59*	*68.87 %*	*Select seq lcl|Query_114124*
Naja naja naja	*214*	*214*	*55%*	*3e-58*	*70.90 %*	*Select seq lcl|Query_114122*

All the sequences shows good score and all the sequences are included in the phylogenetic tree based on neighbour joining (Fig.8) with distance correlation method (Fig. 9) formed by NCBI BLAST.

PHYLOGENETIC TREE BY NCBI BLAST:-

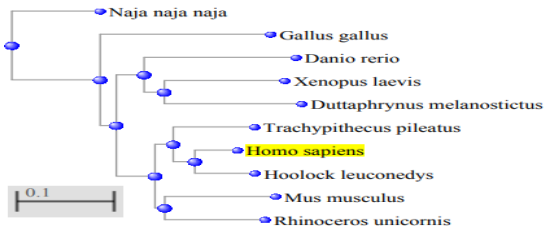


Fig.8—*Phylogenetic tree based on 10 vertebrate series.*

Fig.9 *. Phylogenetic tree among the vertebrate species based on clustal omega.*

Multiple Sequence Alignment Viewer 1.10.0. Regions of similarity are shown by grey and dissimilarity by red colour. The 12S rRNA of Naja naja naja shows the little similarity with others sequences. Region from 481 to 507 base pair shows somewhat similar sequences.

3.4. MEGABLAST for vertebrate series:-

In case of vertebrate series we take the Homo sapiens mitochondrial 12S rRNA as the query sequence and other 9 vertebrate organisms 12S rRNA as our subject sequence. For the highly similar sequence megablast is performed. 8 out of the 9 sequences show highly similar sequences with the query sequence. Only one the 12S rRNA of naja naja naja is not similar at that level and excluded from the megablast result (Table 9).

Table 9:- *Result of Mega Blast*

Organisms	*Max score*	*Total score*	*Query cover*	*e-valu*	*Percent identity*	*Accession*
Hoolock leuconedys	*1240*	*1240*	*100 %*	*0.0*	*90.22 %*	*Select seq lcl|Query_171014*
Trachypithecus pileatus	*920*	*920*	*99%*	*0.0*	*84.41 %*	*Select seq lcl|Query_171015*
Rhinoceros unicornis	*610*	*610*	*99%*	*1e-177*	*78.70 %*	*Select seq lcl|Query_171013*
Mus musculus	*505*	*505*	*84%*	*6e-146*	*78.33 %*	*Select seq lcl|Query_171007*
Danio rerio	*276*	*276*	*60%*	*5e-77*	*75.97 %*	*Select seq lcl|Query_171008*
Xenopus laevis	*202*	*202*	*47%*	*9e-55*	*75.69 %*	*Select seq lcl|Query_171009*
Gallus gallus	*193*	*193*	*57%*	*5e-52*	*73.78 %*	*Select seq lcl|Query_171011*
Duttaphrynus melanostictus	*115*	*115*	*20%*	*1e-28*	*78.17 %*	*Select seq lcl|Query_171012*

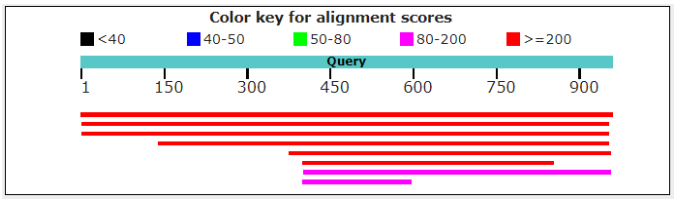


Fig.10. *Distribution of the top 8 BLAST hits on 8 subject sequences.*

The graphic (Fig. 10) shows the highly similar subject sequences with the query sequence. Different colour coded bars indicating the alignment score with the query sequences. Only 2 pink bar represents the 2 organisms with score 80-200 range, in that case these 2 organisms are Gallus gallus which score 193 with the query and other is Duttaphrynus melanostictus with score 115 with the query sequence (Table 10).

Multiple Sequence Alignment :-

Query Id	*Organism Name*	*Length*
lcl|Query_171005	*Homo sapiens*	*954*

Table 10 :- *Multiple subject information*

Query ID	Organisms name	Base pair	Query ID	Organisms name	Base pair
lcl|Query_171007	Mus musculus	955	lcl|Query_171012	Duttaphrynus melanostictus	668
lcl|Query_171008	Danio rerio	952	lcl|Query_171013	Rhinoceros unicornis	971
lcl|Query_171009	Xenopus laevis	819	lcl|Query_171014	Hoolock leuconedys	953
lcl|Query_171011	Gallus gallus	976	lcl|Query_171015	Trachypithecus pileatus	947

The overall similarity and dissimilarity:-

The overall differences in the sequences are shown by red colour , the regions of similarity for grey colour. Naja naja naja's 12S rRNA is not shown by the MSA as it's not highly similar to the other sequences.

One of the conserved sequences in vertebrate series:-

From base pair 444 to 468 within these 25 base pair region only one major nucleotide at position 455 in case of Trachypithecus pileatusis , except that all the nucleotide in each and every position is similar to the query as well as subjects. From these 25 long base pair sequences a good primer can be design with more advance bioinformatics tools.

3.5. Combined results including vertebrate and invertebrate phyla:-

In case of combined phylogenetic tree both the vertebrate and invertebrate series is combined (Table 11) and with the help of NCBI BLAST and CLUSTRAL OMEGA a phylogenetic tree is constructed. Here Query sequence information is

Query Id	*Organism Name*	*Length*
lcl|Query_239659	*Homo sapiens*	*954*

Table 11:- *List of Multiple subject Information*

Query Id	*Organism Name*	*Length*	*Query Id*	*Organism Name*	*Length*
lcl|Query_138963	*Caenorhabditis elegans*	*697*	*lcl|Query_138972*	*Danio rerio*	*952*
lcl|Query_138964	*Drosophila melanogaster*	*786*	*lcl|Query_138973*	*Xenopus laevis*	*819*

lcl|Query_138965	Arbaxia lixula	886	lcl|Query_138974	Naja naja naja	930
lcl|Query_138966	Megascolex sp.	385	lcl|Query_138975	Gallus gallus	976
lcl|Query_138967	Loligo bleekeri	782	lcl|Query_138976	Duttaphrynus melanosticus	668
lcl|Query_138968	Heteractis crispa	788	lcl|Query_138977	Rhinoceros unicornis	971
lcl|Query_138969	Beroe forskalii	376	lcl|Query_138978	Hoolock leuconedys	953
lcl|Query_138970	Oscarella microlobata	1283	lcl|Query_138979	Trachypithecus pileatus	947

The most sequences with good score shown in red colour (Fig.11) , others colour represents the different score for different organisms. For example colour green represents the Arbaxia lixula with alignment score 50-80 with the query sequence. Now with the help of this tabulated information (Table 12 & 13) and by using the NCBI BLAST and CLUSTRAL OMEGA a phylogenetic tree is constructed. But in case of the NCBI BLAST several distantly related organisms with high E value are excluded from the phylogeny and as a consequence only certain phyla which shows good score are included in the phylogeny (Fig 12).

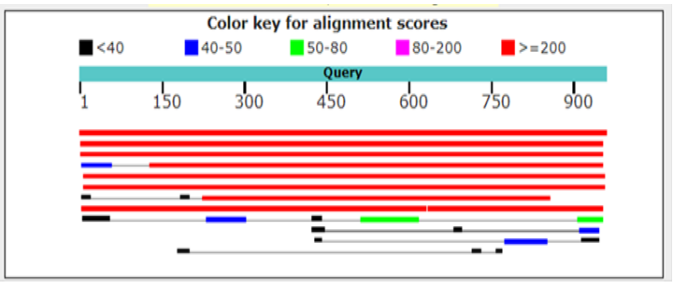


Fig.11 :- *Distribution of the top 34 blast hits on 15 subjects sequences*

Table 12 :- *Sequence producing significant alignment*

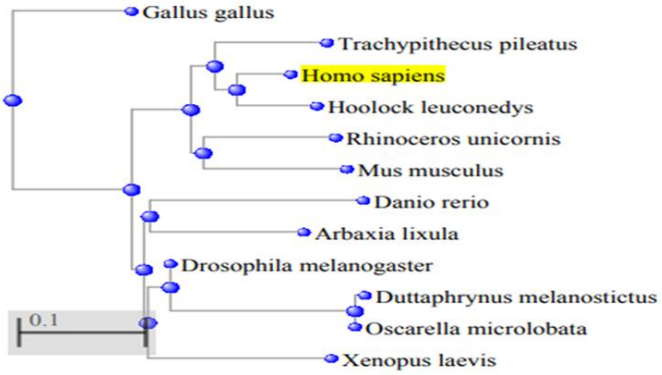
Organisms	Max score	Total score	Query cover	e-valu	Percent identity	Accession
Hoolock leuconedys	1265	1288	100%	0.0	89.64%	Select seq lcl|Query_138978
Trachypithecus pileatus	1003	1068	99%	0.0	84.38%	Select seq lcl|Query_138979
Rhinoceros unicornis	755	755	99%	0.0	78.17%	Select seq lcl|Query_138977
Mus musculus	628	672	91%	0.0	77.74%	Select seq lcl|Query_138971
Danio rerio	398	420	98%	2e-113	70.38%	Select seq lcl|Query_138972
Gallus gallus	394	394	98%	2e-112	70.86%	Select seq lcl|Query_138975
Xenopus laevis	281	339	69%	4e-78	71.32%	Select seq lcl|Query_138973
Duttaphrynus melanosticus	215	237	65%	1e-58	68.87%	Select seq lcl|Query_138976
Naja naja naja	214	214	55%	5e-58	70.90%	Select seq lcl|Query_138974
Arbaxia lixula	59.0	202	30%	2e-11	71.82%	Select seq lcl|Query_138965
Drosophila melanogaster	48.2	104	7%	4e-08	88.89%	Select seq lcl|Query_138964
Oscarella microlobata	44.6	101	12%	5e-07	72.15%	Select seq lcl|Query_138970
Heteractis crispa	29.2	76.0	5%	0.038	86.96%	Select seq lcl|Query_138968
Megascolex sp	21.1	21.1	1%	5.7	100.00%	Select seq lcl|Query_138966
Caenorhabditis elegans	21.1	21.1	1%	5.7	100.00%	Select seq lcl|Query_138963

Table 13:- *Multiple sequence alignment for all the organisms*

Query Id	Organism Name	Length	Query Id	Organism Name	Length
lcl|Query_71169	Caenorhabditis elegans	697	lcl|Query_71178	Danio rerio	952
lcl|Query_71170	Drosophila melanogaster	786	lcl|Query_71179	Xenopus laevis	819
lcl|Query_71171	Arbaxia lixula	886	lcl|Query_71180	Naja naja naja	930
lcl|Query_71172	Megascolex sp	385	lcl|Query_71181	Gallus gallus	976
lcl|Query_71173	Loligo bleekeri	782	lcl|Query_71182	Duttaphrynus melanosticus	668
lcl|Query_71174	Heteractis crispa	788	lcl|Query_71183	Rhinoceros unicornis	971
lcl|Query_71175	Beroe forskalii	376	lcl|Query_71184	Hoolock leuconedys	953
lcl|Query_71176	Oscarella microlobata	1283	lcl|Query_71185	Trachypithecus pileatus	947
lcl|Query_71177	Mus musculus	955			

This problem can be solved with the help of CLUSTRAL OMEGA (Fig.13) which also form the phylogenetic tree with neighbour joining method by without distance corrections. Only 12 organisms form the phylogenetic tree, 3 organisms Heteractis crispa, Megascolex sp and Caenorhabditis elegans excluded from the tree because they show very

poor score and only little 5% to 1% is covered with the query sequences.



.

Fig.12 *Phylogenetic tree devoloped by NCBI BLAST.* This is a Neighbour-joining tree without distance corrections.

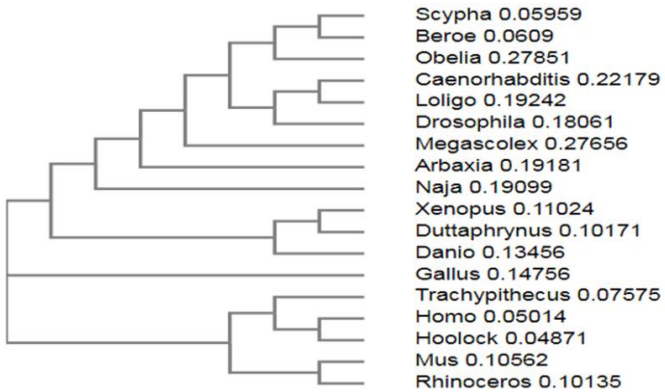


Fig. 13. *Phylogenetic Tree by Clustral Omega*

4. CONCLUSION:-

Total 18 species from vertebrate and invertebrate series is selected for this study. With the help of NCBI data base at first 12S-rRNA of these organisms were downloaded and then NCBI BLAST programme is performed, to know the sequences divergence, similarity etc. at first a table is formed to represented all the organisms with their respective sequences and sequence related information. In the initial stages of this study at first the organisms were divided into two different groups: vertebrate and invertebrate. Then performed the NCBI BLAST and MEGABLAST separately for both the phyla. With the help of MULTIPLE SEQUENCE ALINGMENT tool the sequences were also analyzed and the region of similarity-dissimilarity was identified. In case of invertebrate series the 8 organisms was not well response to the MEGABLAST, indicating that they were not highly similar to each other on the basis of 12S rRNA. Only 2 species from different 2 phyla were well respond to the blast. But in case of vertebrate series except Naja naja naja the other eight organisms were well respond to each other, indicating that they all share an evolutionary conserved sequences with each other. These conserved sequences can be utilized for designing a good primer. For the determination of somewhat similar sequences normal BLAST programme was performed, and the sequences which share somehow similar sequences can be determined. The results of blast and megablast both indicating the somehow similar to very similar sequences, and with the proper bioinformatics tools phylogenetic tree was constructed in every possible cases. From these phylogenetic trees the evolutionary of the species can be determined to some extent. As the NCBI BLAST construct the phylogenetic tree on the basis of Neighbour-joining method with distance corrections, the organisms which has the too much sequence divergence with query sequences are simply excluded from the tree, so only organisms with good alignment score are producing the tree. So in some case the total numbers of subject and query sequences are not similar to the number of organisms represented by the tree. So at the same time clustral omega is also performed which producing tree without distance correction and always forms a tree with all the subjects as well as query sequences. This kind of study can indicate the evolutionary divergence on the basis of nucleic acid sequencing which is one of the best current approaches for study molecular diversity within or between the species. The evolutionary conserved domain can be used for the universal primer designing, by which species specific primers can also be formed. The bioinformatics tools can also be used for the determination of the nucleotide to encoded amino acids, from which the structure of protein, with their respective function can be determined. Mutation within the specific region of the gene can also affect the encoded protein, in that cases sequencing can help us to determine the site of mutation.

5 ACKNOWLEDGEMENTS: -

The authors express sincere gratitude to the officer-in-charge for providing laboratory facilities for the computational work. Thanks are extended to the Librarian of A B N Seal College, Cooch Behar for extending extra mural support.

6 REFERENCES:-

[1] Hodgman, C, French, A and Westhead D" Bios instant notes in Bioinformatics".(ISBN 978-0-4153-9494-9) Garland Science, USA. First Indian Edition,19-28,107-173, 2015.

[2] Attwood, T. K , Parry-smith, D. J and Phukan. S Introduction to bioinformatics (ISBN 978-81-7758-641-1) Dorling Kindersley (India) Pvt. Ltd. First Edition, 1-218, 2011.

[3] Siddappa,C,Saini,M.,Das, A, Doreswamy,R, Sharma,A.K, Gupta.P.K [ed.] Malayannan B. Subramaniam.Sequence Characterization of Mitochondrial 12S rRNA Gene in Mouse Deer (Moschiola indica) for PCR-RFLP Based Species Identification. 20 November 2013, s.l. : Hindawi Publishing Corporation, 2013, Vol. Molecular Biology International Volume 2013. 783-925, 2013.

[4] Li Yang Zongqing Tan, Daren Wang, Ling Xue Min-xin Guan, Taosheng Huang, Ronghua Li."Species identification through mitochondrial rRNA genetic analysis". NATURE. | 4 : 4089 |, 2015. DOI: 10.1038/srep04089.

[5] Giribet G, Dunn C W, Edgecombe G.D, greg w. Rouse, G,W "A modern look at the Animal Tree of Life" Zootaxa 1668, 61–79, 2007
[6] Bourlat S.J, Nielsen C, Economou, A.D, Maximilian J. Telford, J. M "Testing the new animal phylogeny: A phylum level molecular analysis of the animal kingdom" Molecular Phylogenetics and Evolution 49, 23–31, 2008.
[7] E - Resource:-
[8] (https://www.ebi.ac.uk/Tools/psa/) accessed on 13th October 2019.
[9] (https://blast.ncbi.nlm.nih.gov/) accessed on 13th October 2019.
[10] (https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome) accessed on 13th October 2019
[11] (https://www.ebi.ac.uk/Tools/msa/) accessed on 13th October 2019.
[12] (https://www.ebi.ac.uk/Tools/msa/clustalo/) accessed on 13th October 2019.