International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2016): 79.57 | Impact Factor (2017): 7.296

Comparative Study and Tree Construction of Different Organism Based on Histone Family

Akhilesh Lokhande[1], **Dinesh Maske**[2], **Yusuf Talib**[3]

[1]Institute of Bioscience and Technology, MGM campus, Aurangabad. (M.S) 431001

[2]MGMs G.Y.Pathrikar College of CS and IT, MGM Campus, Aurangabad.(M.S)431001
Department of Biotechnology, Dr.Rafiq Zakaria Campus,

[3]Maulana Azad College of Arts, Science and Commerce, Aurangabad. (M.S) 431001

Abstract: ***Phylogenetic is one of the branches to study evolutionary study of living organisms, the classification of organism based on basis of similarity. Bioinformatics plays a key role to modulate and enhance the study of sequence similarity and provides the detailed study of each and every sequence and their evolutionary study using the concepts of Phylogenetic tree analysis, the concept help researcher to understand the importance of divergence, time factor, and pedigree of sample sequences. Histones are highly alkaline proteins found in eukaryotic cell nuclei that package and order the DNA into structural units called nucleosomes Genes encoding histone variants are usually not clustered, have introns and their mRNAs are regulated with polyA tails. Five major families of histones exist:H1/H5,H2A,H2B,H3, andH4. Histones H2A, H2B, H3 and H4 are known as the core histones, while histones H1/H5 are known as the linker histones.***

Keywords: Phylogenetic, Alignment, Histone, Phylogram, Cladogram

1. Introduction

Histone are highly alkaline proteins found in eukaryotic cell nuclei that package and order the DNA into structural units called nucleosomes.They are the chief protein components of chromatin, acting as spools around which DNA winds, and playing a role in gene regulation. Without histones, the unwound DNA in chromosomes would be very long (a length to width ratio of more than 10 million to 1 in human DNA). Histones are subdivided into canonical replication-dependent histones that are expressed during the S-phase of cell cycle and replication-independent histone variants, expressed during the whole cell cycle. In animals, genes encoding canonical histones are typically clustered along the chromosome, lack intron and use a stem loop structure at the 3' end instead of a poly-A tail. Genes encoding histone variants are usually not clustered have introns and their mRNAs are regulated with poly-A tails. Five major families of histone exist: H1/H5, H2A, H2B, H3, and H4. Histones H2A, H2B, H3 and H4 are known as the core histones, while histones H1/H5 are known as the linker histones.

Our aim to study the similarity between different organisms based on the marker (Histone family) using evolutionary studies. Phylogenetic inference is the process of developing hypothesis about the evolutionary relatedness of organisms based on their observable characteristics. When Linnaeus developed the system of classification into kingdom, phyla genera, and species, the early biologists sorted living things into a symbolic tree of life. This tree based representation of the relationships among species is a phylogenetic tree.
Evolutionary distance between pairs of sequence, relative to other sequences in an input data set, is one way to assign branch length. While a phylogeny of species generally has as root, assuming that all species have a specific common ancestor, a phylogenetic tree derived from sequence data may be rooted or unrooted.

2. Programs

Different protocols and methods are available by which information of common ancestor can be drawn for that firstly target sequence is obtained in specific fasta file format (developed by W. Pearson) form biological database, obtained sequence here Histone families (H1/H5) these sequence has processed for sequence alignment, similar sequence has been shortlisted on the basis of Expectation Value (E-Value) (Table 1), then those shortlisted sequence has been performed by using **CLUSTALX2** drawn in the form of fasta for tree construction program using ClustalX (Provides output in .dnd, .phy), Tree view and Phylip. Shown in figure 1 and 2.

Table represents all species in phylogenetic tree:

Organism name	Scientific name	Accession number
Human	Homo sapiens	AAC61625.1
Cow	Bos Taurus	NP_001030449.1
Dog	Canis lupus familiaris	NP_001274021.1
Donkey	Equus asinus	XP_014707060.1
Horse	Equus caballus	NP_001243880.1
Tobacco	Nicotiana tabacum	NP_001312320.1
Oats	Avena sativa	AAA32718.1
Rat	Rattus norvegicus	NP_072169.1
Wheat	Triticum aestivum	AAB00193.1
Pig	Sus scrofa	NP_001116594.1
Rice	Oryza sativa	AAA33907.1
Cabbage	Brassica oleracea var. capitata	BAF63143.1
Barley	Hordeum vulgare subsp. vulgare	AEK21393.1
Pea	Pisum sativum	BAM74171.1
Maize	Zea mays subsp. Mays	AAQ62067.1

3. Phylogenetic tree construction

Those obtained output from clustalX in .dnd and .aln files are used as a input in phylip for Phylogenetic Tree Construction (rooted and unrooted form) **fig.1, fig.2, fig.3** and **fig.4**. Those outputs show the position of obtained sequence in evolutionary tree.

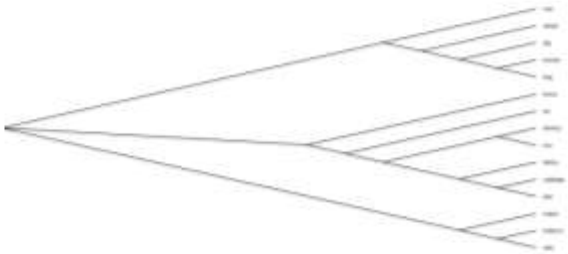

Figure 1: slanted cladogram

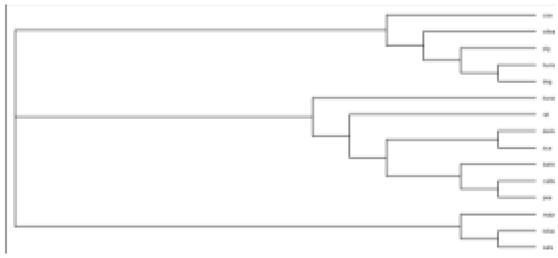

Figure 2: rectangular cladogram

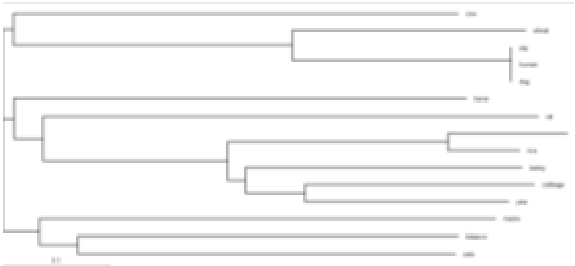

Figure 3: phylogram

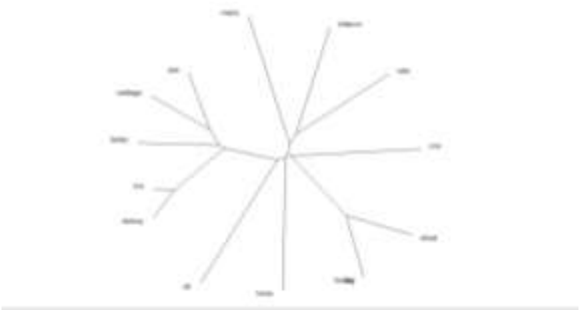

Figure 4: unrooted

4. Materials and Methods

Phylip:
PHYLIP (the PHYLogeny Inference Package) is a package of programs for inferring phylogenies (evolutionary trees). Methods that are available in the package include parsimony, Distance matrix, and likelihood methods, including bootstrapping and consensus trees. Data types that can be handled include molecular sequences, gene frequencies, restriction sites and fragments, distance matrices, and discrete characters. The data is read into the program from a text file, which the user can prepare using any word processor or text editor. Some sequence analysis programs such as the ClustalX2 alignment program can write data files in the PHYLIP format.

Clustalx2:
Clustalx2 is a new windows interface for clustalW multiple sequence alignment program. It provides an integrated environment for performing multiple sequence and profile alignment and analysing the results. The sequence alignment is displayed in a windows on the screen. The pull-down menus at the top of the window allow you to select all the option required for traditional multiple sequence and profile alignment.

TreeView:
TreeView is a program which is used to visualize a phylogenetic tree (**slanted cladogram, rectangular cladogram, phylogram, unrooted**), this is one of the application which is used to visualize a phylogenetic tree and able to understand different node, ancestor.

5. Conclusion

Phylogenetic trees contain a lot of information about the inferred evolutionary relationships between different organisms/species**. In fig.1, fig.2, fig.3** and **fig.4** the horizontal dimension gives the amount of genetic change. The horizontal lines are branches & are represents. Evolutionary lineages changing over time. The longer the branch in the horizontal dimension, longer the amount of change.

For our study Histone has been utilized as a genetic marker for similarity and various tree constructions using the applications like PHYLIP, TREEVIEW. Clustal X2 gave alignment, distance matrix, to generate the phylogram, cladogram. Samples of histone from various organisms extracted, based on they show the maximum similarity after tree construction. Four branches obtained after construction one shows large cluster of mammals (Human, pig, horse, Donkey, Dog, cow), small group belongs to other species (rice, wheat, barley, pea, maize) another group belongs some animals (rat), which concludes Histone (marker for this experiment) having similarity with different organism possess same fragment of sequence. The molecular clock hypothesis states that Protein sequences evolve at a rate that is relatively constant over time and among different organisms.

References

[1] Hodge T, & M. J. T. V. Cope. 2000. a myosin family tree journal of cell science.
[2] Woese C. R. 1998.
[3] Maher B. A. 2002 Uprooting tree of life.
[4] Felsenstein J(2004) Inferring phylogenesissinauer associated sunderland.
[5] Penny, D, Hendy, M.D & M.A. Steal 1992.
[6] Hodge T, Cope MJ (2002) A myosin family tree.
[7] Essential Bioinformatics (Jin Xiong)
[8] David Mount: Bioinformatics.