Indian J. Applied & Pure Bio. Vol. 37(1), 38-46 (2022). **ISSN: 0970-2091**
A web of Science Journal

Phylogenetic network Analysis of CDC20 Genes

Parul Johri

Amity Institute of Biotechnology, Amity University Uttar Pradesh,
Lucknow Campus, Malhaur, Gomti Nagar Extension, Lucknow-226010 (India)
pjohri@lko.amity.edu; +91 9838144680

Abstract

Phylogenetic analysis gives an insight into evolution and determines the historical relationships among species. A valid establishment of phylogenetic tree allows us to identify the lineage of group of organisms in which a specific character appeared or disappeared due to mutation. The mechanism of evolution of any specific character of interest can be understood through phylogenetics. CDC20 gene, encodes CDC20 protein which plays a significant role as the regulator of cell division. This gene is present throughout the eukaryotic life forms from Kingdom Protista to Animalia *i.e.*, from lowest eukaryote to highest eukaryote.

CDC20 genes of various eukaryotes were retrieved from online databases and were subjected to multiple sequence analysis using ClustalX and further for phylogenetic analysis using Phylip software. The phylogenetic tree hence generated was analyzed for clads and clustering.

Three main clads were identified with some more intermittent clades. The Phylogenetic tree showed that CDC20 gene is the primitive gene showing its existence from early eukaryotes *i.e*. Kingdom Protista to recently evolved eukaryotes Kingdom Animalia. The results showed phylogenetic position of CDC20 gene in diverse levels of cataloging in different Kingdoms. Some clusters showed up to 83-85% divergence among organisms belonging to same or different Kingdom. These annotations show that CDC20 gene is conserved during the course of evolution and that the species in the clades are close kinsfolks to each other.

The phylogenetic tree raised using PHYLIP software shows that the kingdom animalia has evolved the latest during the course of evolution of CDC20 gene supporting the view of Five Kingdom Classification by R.H. Whittaker.

Phylogenetics is the science to study the evolutionary relationship amongst various organisms, genes or protein sequences. Mutational changes in genes are the primary cause of evolution. These evolutionary relationships are predicted by the construction of phylogenetic trees, which links the individuals or group of organisms[1]. Earlier, phylogenetics dealt essentially with physical or morphological features like size, color, number of legs, etc. Recent phylogeny uses information extracted from genetic material, mainly from DNA and protein sequences. The characters used are usually the DNA or amino acid residues, after aligning quite a few such sequences, and using only blocks which were conserved in all the examined species. During evolution, it is very frequent for a gene to get duplicated. The copies continue to evolve independently, resulting in two (or more) similar instances of the same gene along the genome of a species[2]. Therefore, matching genes in different species get differentiated between orthologous and paralogous genes. Orthologs are the genes that went through a speciation event, they are connected directly, and not through a duplication. Orthologs, or orthologous genes, are genes in different species that are similar to each other because they originated from a common ancestor. Paralogs are the genes that were separated by the gene duplication event. Paralogs typically have the same or similar function, but sometimes do not: due to lack of the original selective pressure upon one copy of the duplicated gene, this replica is free to mutate and acquire new functions.

CDC20 gene, encodes CDC20 protein which plays a significant role as the regulator of cell division. This gene is present throughout the eukaryotic life forms from Kingdom Protista to Animalia *i.e.*, from lowest eukaryote to highest eukaryote. The cytogenic location of Cdc20 gene on chromosome in humans is 1p34.2. CDC20 protein mainly activates Anaphase Promoting Complex (APC). It is also necessary for two microtubule dependent processes viz., nuclear movement preceding anaphase and chromosome separation. Until all the chromosomes are aligned at the metaphase plate, in the M phase of cell cycle, the spindle assembly checkpoint mechanism delays anaphase initiation. Activation of the APC by binding of CDC20 and CDH1 is mandatory for exit from mitosis. Furthermore, APC has been concerned as a target for the checkpoint intervention. Mitotic arrest deficient (MAD) protein is a key component of this checkpoint system inhibits APC. Over expression of CDC20 gene is established to develop tumor in human beings. As a result, targeting CDC20 gene with its inhibitors could be a novel strategy for the cancer treatment. For instance, p^{53} protein is the indirect regulator of CDC20 and thus inhibits tumor cell growth. This might be a good potential therapeutic target for a wide range of human cancer. Thus, the current study was carried out to depict insights concerning the relationship between the CDC20 genes of various eukaryotes that belong to various taxonomic groups[3]. Considering the literature survey, the CDC20 gene consists a vital oncogenic function in the development and succession of human cancers, which when inhibited by certain inhibitors, be capable of playing as a novel therapeutic target for cancer. Such results assist the researchers to represent insights for gene therapy and the

conserved regions of Cdc20 gene of lower eukaryotes can be supportive for understanding and implementing necessary changes for the development of therapeutic capability of CDC20 gene and its inhibitors. In Phylo Oncology, the use of phylogenetics is a prominent systems biology approach, from combining subsets of cancer samples to tracing subclonal evolution throughout cancer development and metastasis[4]. Well-developed phylogenetic applications offer fast, robust approaches to evaluate high-dimensional, diverse cancer data sets[5].

Data Collection :

20 nucleotide sequences of CDC20 genes of various eukaryotes that belong to Kingdom Protista, Fungi, Plantae and Animalia were retrieved from the NCBI GenBank database. All these sequences were downloaded in FASTA format and were compiled into a single text file for performing multiple sequence alignment. Table 1 contains the details of all the sequences analyzed[6,7].

Table-1. Accession numbers of CDC20 genes of 20 nucleotide sequences retrieved from GenBank database

S. No.	Accession Number	Scientific Name	Kingdom
1	U66069.1	*Tritichomonas foetus*	Protista
2	AQ848502.1	*Leishmania major*	Protista
3	CCKQ01004474.1	*Stylonychia lemnae*	Protista
4	NBNE01003783.1	*Phytopthora megakarya*	Protista
5	ABRE01013192.1	*Monilophthora perniciosa*	Fungi
6	LTAI01001460.1	*Hepatospora eriocheir*	Fungi
7	LGUB01000472.1	*Pseudoloma neurophilia*	Fungi
8	X59428.1	*Saccharomyces cerevisiae*	Fungi
9	U77983.1	*Schizosaccharomyces pombe*	Fungi
10	AY675107.1	*Ostreococcus tauri*	Plantae
11	KD495891.1	*Triticum urartu*	Plantae
12	AF029264.1	*Arabidopsis thaliana*	Plantae
13	JRRC01041684.1	*Gossypium arboretum*	Plantae
14	ASHM01104304.1	*Trifolium pretense*	Plantae
15	LK042044.1	*Brassica napus*	Plantae
16	MF432992.1	*Diachasma muliebre*	Animalia
17	AVOS01078843.1	*Chaetura pelagica*	Animalia
18	DQ473545.1	*Homo sapiens*	Animalia
19	A4411083.1	*Mus musculus*	Animalia
20	A4411082.1	*Pan troglodytes*	Animalia

Multiple sequence Alignment :

Clustal X 2.1 was used for performing Multiple Sequence Alignment, which is an offline tool that performs optimum alignment for multiple nucleotide and protein sequences[8,9]. The required outputs for further analysis and tree construction were a multiple sequence alignment in alignment format and phylip format, respectively (Figure 1)[10].

Phylogenetic Tree construction :

The phylogenetic tree was constructed using PHYLIP Software (version 3.695). The MSA data was analyzed using two programs of PHYLIP software[11,12]:

1. protdist: It calculates the distance among the aligned sequences. It takes the phylip output format as the infile (Figure 2).

2. neighbor: It constructs the phylogenetic tree using neighbor-joining method. The outfile of the protdist program is the infile of the neighbor. The outfile of this program is generated in tree extension for the visualization of the phylogenetic tree (Figure 3).

Visualizing the phylogenetic tree :

The phylogenetic tree resulting from PHYLIP software was visualized in Fig. Tree v1.4.4 (Figure 4).

All the CDC20 genes selected from 20 eukaryotic species belong to various Kingdoms. The ClustalX software gave the reports of Multiple Sequence Alignment which shows numerous gaps among the species. The output file was generated in phylip format and was used as an input file for the PHYLIP programs, for further phylogenetic analysis and construction of evolutionary tree. The phylogenetic analysis demonstrates that CDC20 gene has diverged extremely during the course of its evolution. The PHYLIP software was used to develop phylogenetic tree and this can be divided into three clades, A, B and C. These distinct clades are further divided into sub-clades.

Clade A - This clade consists of CDC20 sequences of Plantae, Protista and Fungi Kingdoms. The sequences diverged into Plantae are kept in one group and sequences diverging into Fungi and Protista are kept in another. These subclades contain sequences as shown in the Tables-2 and 3.

Table-2. Subclade A.1

S.No.	Accession numbers	Scientific name	Kingdom	% Divergence
1	KD495891.1	*Triticum urartu*	Plantae	57.995
2	JRRC01041684.1	*Gossypium arboreum*	Plantae	46.116
3	LK042044.1	*Brassica napus*	Plantae	70.061

This subclade shows a divergence among the sequences between 46 to 70%.

Table-3. Subclade A.2

S.No.	Accession numbers	Scientific names	Kingdom	% Divergence
1	NBNE01003783.1	*Phytophthora megakarya*	Protista	85.872
2	CCKQ01004474.1	*Stylonychia lemnae*	Protista	65.126
3	LGUB01000472.1	*Pseudoloma neurophilia*	Fungi	53.958

The second subclade shows the sequence divergence between 53 to 85%.

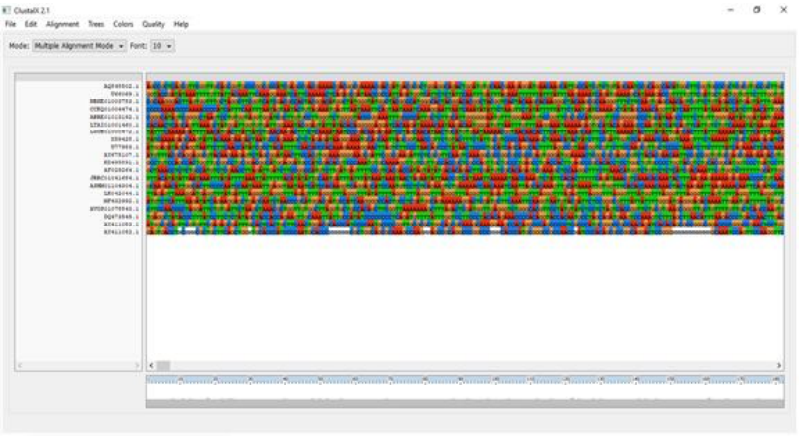

Figure 1: Complete alignment output using ClustalX 2.1

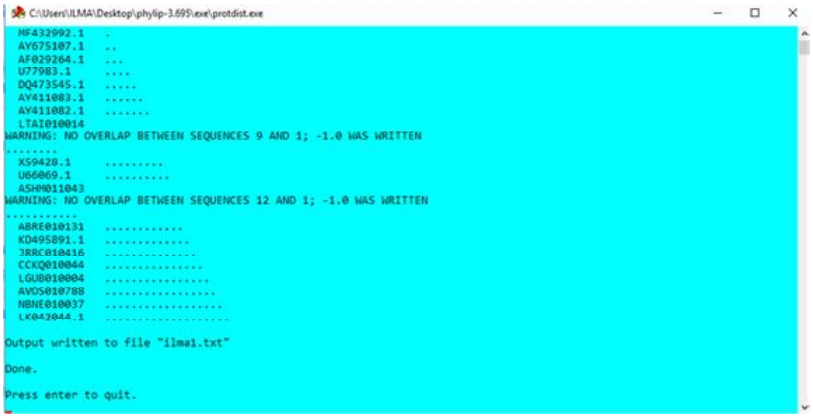

Figure 2: protdist program to calculate the distance among the aligned sequences

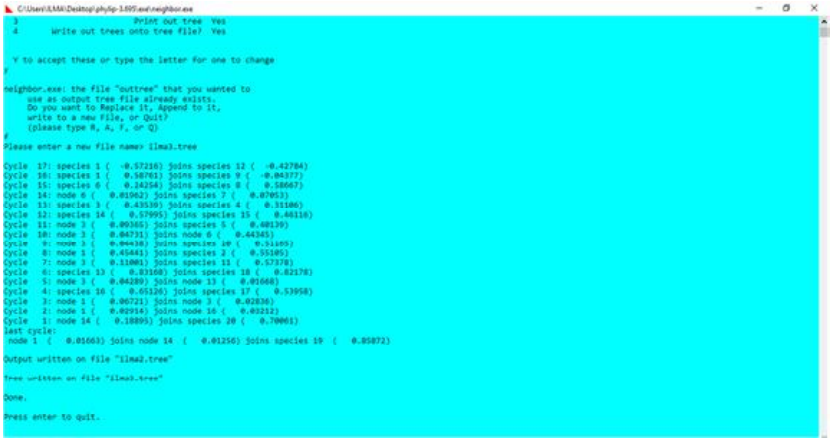

Figure 3: neighbor program using neighbor-joining method to construct phylogenetic tree.

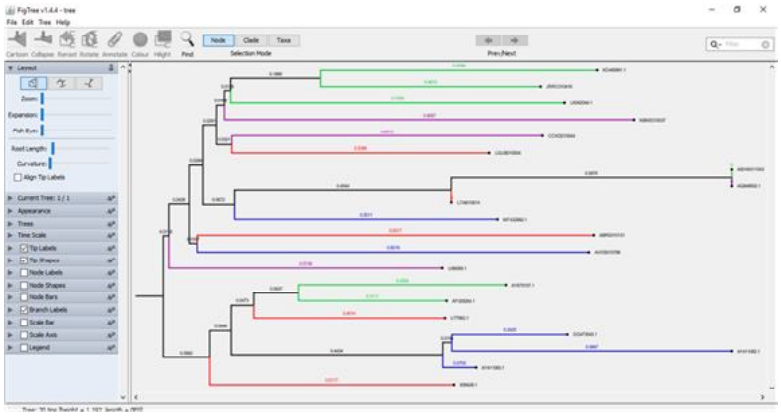

Figure 4: Phylogenetic tree visualization using PHYLIP

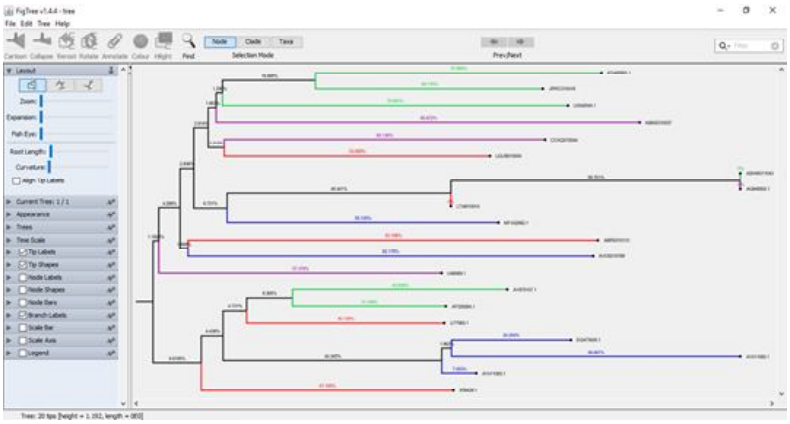

Figure 5: Phylogenetic tree visualization using FigTree showing percent divergence of CDC20 gene

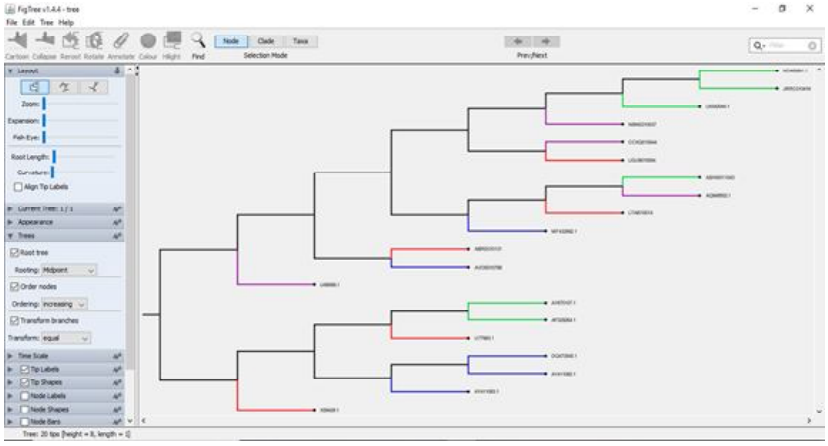

Figure 6: Phylogenetic tree showing evolution earliest to latest

Clade B - This clade consists of CDC20 sequences of Plantae, Protista, Fungi and Animalia Kingdoms. This clade has a distinct out group *Tritichomonas foetus* belonging to Kingdom Protista. All the sequences have diverged differently and are divided in subclades which are shown in Tables-4 and 5.

Table-4. Subclade B.1

S.No.	Accession numbers	Scientific name	Kingdom	% Divergence
1	ASHM01104304.1	*Trifolium pratense*	Plantae	0.000
2	AQ848502.1	*Leishmania major*	Protista	0.000
3	LTAI01001460.1	*Hepatospora eriocheir*	Fungi	0.000
4	MF432992.1	*Diachasma muliebre*	Animalia	55.105

This subclade shows the divergence among the sequences between 0 to 55%.

Table-5. Subclade B.2

S.No.	Accession numbers	Scientific name	Kingdom	% Divergence
1	ABRE01013192.1	*Monilopthora perniciosa*	Fungi	83.168
2	AVOS01078843.1	*Chaetura pelagica*	Animalia	82.178
3	U66069.1	*Tritichomonas foetus*	Protista	57.378

This subclade shows the divergence among organisms between 57 to 83%.

Clade C - This clade consists of CDC20 sequences of Plantae, Fungi and Animalia Kingdoms. The sequences diverging into Plantae and Fungi are kept in one group and the sequences diverging into Animalia and Fungi are kept in another group. This clade has a distinct out group Saccharomyces cerevisiae belonging Kingdom Fungi. The subclades are shown in Table 6 and Table 7.

Table-6. Subclade C.1

S.No.	Accession numbers	Scientificname	Kingdom	% Divergence
1	AY675107.1	*Ostreococcus tauri*	Plantae	43.539
2	AF029264.1	*Arabidopsis thaliana*	Plantae	31.106
3	U77983.1	*Schizosaccharomyces pombe*	Fungi	40.139

The percentage of divergence among organisms in this subclade is between 31 to 43%.

Table-7. Subclade C.2

S.No.	Accession numbers	Scientific name	Kingdom	% Divergence
1	DQ473545.1	*Homo sapiens*	Animalia	24.254
2	AY411082.1	*Pan troglodytes*	Animalia	58.667
3	AY411083.1	*Mus musculus*	Animalia	7.053
4	X59428.1	*Saccharomyces cerevisiae*	Fungi	51.165

In this subclade, the sequences show a percentage divergence between 7 to 58%. The sequences of CDC20 gene showed some interesting results after the construction of phylogenetic tree. The sequences of CDC20 of Kingdom Plantae and Kingdom Animalia segregated specifically into clades over the dendogram. Also, the sequences are clustered according to their phylogenetic relations that is, the sequences belonging to the same order. There were some exceptional sequences that extended as the out group belonging to Kingdoms Protista and Fungi. The similarities among the organisms are shown which are placed nearby to each other. Thus, the phylogenetic tree constructed using PHYLIP software shows that the Kingdom Animalia has evolved the latest during the course of evolution of CDC20 gene supporting the view of Five Kingdom Classification by R.H. Whittaker (Figure 5 and Figure 6)[13,14].

The phylogenetic study was carried out using PHYLIP software (version 3.695) which showed some interesting results about the evolution of CDC20 gene throughout the various life forms in different eukaryotes. The Kingdom Plantae and Kingdom Animalia showed a segregated diversion and thus branched and clustered separately in the phylogenetic tree. The Fungi showed a close relation with all the Kingdoms. Similarly, there are such certain patterns of the divergence and branching of CDC20 genes in the resulting phylogenetic tree. Based on the research survey, CDC20 possess a crucial oncogenic property which may lead to the development and progression of tumor ultimately causing cancer in human beings. This can play as a novel therapeutic agent, if inhibited by certain inhibitors. Such analysis help the researchers to represent insights for gene therapy and the conserved regions of CDC20 gene in lower eukaryotes can be useful for understanding and implementing required changes for the development of therapeutic ability of CDC20 gene and its inhibitors[15,16].

In this study, evolutionary relationship of CDC20 among 20 eukaryotic species illustrated the importance of identifying a variety of unique features such as variation and conserved regions. The Phylogenetic tree showed that CDC20 gene is the primitive showing its presence from early eukaryotes i.e. Kingdom Protista to recently evolved eukaryotes Kingdom Animalia. The results showed phylogenetic position of CDC20 gene in different levels of classification in different Kingdoms. Some clusters showed up to 83-85% divergence among organisms belonging to same or different Kingdom. These observations show that CDC20 gene is conserved during the course of evolution and that the species in the clades are close relatives to each other. This provides the researchers opportunity to explore insights among the sequences to initiate research work by comparative approach. Based on research works, the CDC20 gene plays a vital oncogenic role in the growth and progression of human cancers. With the use of PhyloOncology, a promising subject in phylogenetics, additional analysis and in-depth exploration can be performed for evolutionary analysis of cancer amongst the eukaryotes.

References :

1. Chi S.W., A. Ayeda, and C.H. Arrowsmith (1999). *EMBO J. 18:* 4438-4445.
2. De Laurenzi V., A. Costanzo, D. Barcaroli, A. Terrinoni, M. Falco, M. Annicchiarico-Petruzzelli, M. Levrero, and G. Melino (1998). *J Exp. Med. 188:* 1763-1768.

3. De Laurenzi V.D., M.V. Catani, A. Terrinoni, M. Corazzari, G. Melino, A. Costanzo, M. Levrero, and R.A. Knight (1999). *Cell Death Differ. 6:* 389-390.
4. Doyle J.J., and M.A. Luckow (2003). *Plant Physiol. 131:* 900-910.
5. Ikawa S., A. Nakagawara, and Y. Ikawa (1999). *Cell Death Differ. 6:* 1154-1161.
6. Kaghad M., H. Bonnet, A. Yang, L. Creacier, J.C. Biscan, A. Valent, A. Minty, P. Chalon, J.M. Lellas, X. Dumont, P. Ferrara, F. McKeon, D. Caput (1997). *Cell 90:* 809-819.
7. Keyes W. M., Y. Wu, H. Vogel, X. Guo, S.W. Lowe, and A.A. Mills (2005). *Genes Dev. 19:* 1986-1999.
8. Melino G., F. Bernassola, M. Ranalli, K. Yee, W.X. Zong, M. Corazzari, R.A. Knight, D.R. Green, C. Thompson and K.H. Vousden (2004). *J. Biol. Chem. 279:* 8076-8083.
9. Mills A.A., B. Zheng, X.J. Wang, H. Vogel, D.R. Roop, and A. Bradley (1999). *Nature 398:* 708-713.
10. Nomoto S., N. Haruki, M. Kondo, H. Konishi, T. Takahashi, T. Takahashi and T. Takahashi (1998). *Cancer Res. 58:* 1380-1383.
11. Pavlopoulos G.A., G.S. Theodoros, B.S. Adriano, and S Reinhard (2010). *Bio Data Min. 3*(1): 1.
12. Pintus S.S., and V.A. Ivanisenko (2006) phylogenetic analysis of the p53 AND p63/p73 gene families. Comparative and evolutionary genomics and proteomics.
13. Pozniak C.D., S. Radinovic, A. Yang, F. McKeon, D.R. Kaplan, and F. D. Miller (2000). *Science 289:* 304-306.
14. Pryer K.M., H. Schneider, E.A. Zimmer, and J.A. Banks (2002) *Trends Plant Sci. 7:* 550-554.
15. Serber Z., H.C. Lai, A. Yang, H.D. Ou, M.S. Sigal, A.E. Kelly, B.D. Darimont, P.H.G. Duijf, van H. Bokhoven, and F. McKeon, *et al*. (2002). *Mol. Cell Biol. 22:* 8601-8611.
16. Soltis D.E., and P.S. Soltis (2003). *Plant Physiol. 132:* 1790-1800.