



A COMPARATIVE ACCOUNT ON *IN SILICO* PREDICTION OF EVOLUTIONARY DIVERGENCE IN CDC20 GENE OF VARIOUS EUKARYOTES, USING MESQUITE, MEGA AND PHYLIP

Rokkam Prasuja¹, Mavuduri Jagannath^{2*} and Rupesh S. Badere³

¹Trainee, DBT BIF Programme, Department of Marine Living Resources, Andhra University, Visakhapatnam.

²Research Assistant, DBT BIF Programme, Department of Marine Living Resources, Andhra University, Visakhapatnam.

³Assistant Professor, Department of Botany, RTM Nagpur University, Nagpur-440033.

***Corresponding Author: Mavuduri Jagannath**

Research Assistant, DBT BIF Programme, Department of Marine Living Resources, Andhra University, Visakhapatnam.

Article Received on 10/12/2017

Article Revised on 31/12/2017

Article Accepted on 20/01/2018

ABSTRACT

Cell-division cycle 20 gene (Cdc20), which encodes Cdc20 protein is an important regulator of cell division. This gene is present from Kingdom Protista to Animalia i.e., from lowest eukaryote to highest eukaryote with many aliases like p55Cdc, fizzy, etc. The current study was undertaken to evaluate the phylogenetic divergence of Cdc20 genes selected randomly from 47 eukaryotic organisms. These gene sequences were retrieved from Genbank database of NCBI. This was followed by Multiple sequence alignment performed with Clustal X software. The phylogenetic trees were constructed using PHYLIP, MEGA and MESQUITE. The results show that Cdc20 gene was present among the lowest to highest eukaryotes and also reflects a good sequence similarity amongst the conserved regions of all Cdc20 genes, within the organisms of same order. Certain groups of the Phylogram also showed more than 90% sequence similarity across the conserved regions. An attempt has also been made to compare the trees built from the three programs. The literature review also suggests that the Cdc20 gene plays a critical oncogenic role in the development and progression of cancer. Therefore, using the approaches of PhyloOncology, the evolutionary relationship of this gene can be predicted based on cancer development and metastasis.

KEYWORDS: Cdc20 genes, Phylogenetics, PHYLIP, MEGA, MESQUITE, Cancer, PhyloOncology.

1. INTRODUCTION

Cdc20 gene, which encodes Cdc20 protein plays an important role as regulator of cell division. This gene is present throughout the eukaryotic life forms i.e., from lowest eukaryote to highest eukaryote. Cdc20 has many aliases like p55Cdc, fizzy, Cdc20A, bA276H19.3, SPBC25H2.13c, C87100. The cytogenic location of this gene on chromosome in humans is 1p34.2. Cdc20 protein mainly activates Anaphase Promoting Complex (APC) but it is also required for two microtubule-dependent processes viz., nuclear movement prior to anaphase and chromosome separation.^[1,2] Hartwell et.al synthesized cell-division cycle mutants that couldn't complete major events in the cell cycle of *Saccharomyces cerevisiae*.^[3] Mutants that could not complete mitosis were unable to enter anaphase.^[4] The amino acid sequences of p55Cdc, a regulatory protein in mammalian cells, contain seven repeats which share the highest degree of homology with *Saccharomyces cerevisiae* Cdc20 and Cdc4 proteins. Cdc20 and Cdc4 have been proposed to be involved in the formation of a

functional bipolar mitotic spindle in yeast cells.^[5,6] For chromosome segregation in *S. cerevisiae* Cdc20 gene is required whose product contains WD40 repeats.^[7-8] APC is an ubiquitin-dependent proteolytic system which degrades Cyclin B during mitosis. Mitotic specificity of Cyclin B ubiquitination is determined by a 20S complex that contains homologs of yeast Cdc16 and Cdc27.^[10,11] APC is a large multimeric cullin-RING ligase and controls segregation of chromatids, cytokinesis and establishment of G1 phase of cell cycle.^[12] Until all chromosomes are aligned at the metaphase plate, the spindle assembly checkpoint mechanism delays anaphase initiation. Activation of the APC by binding of Cdc20 and CDH1 is required for exit from mitosis. Moreover, APC has been implicated as a target for the checkpoint intervention.^[13] Mitotic arrest deficient protein (MAD) is a key component of this checkpoint system, which inhibits APC by balancing its activator P⁵⁵Cdc/Hc20.^[14] Mutagenesis and NMR titration experiments have shown that a C-terminal flexible region of MAD is required for binding to Cdc20.^[15]

Overexpression of Cdc20 is proven to develop tumor in humans. Therefore, targeting Cdc20 with its inhibitors could be a novel strategy for the cancer treatment.^[16] For example, p53 protein indirectly regulates Cdc20 and thus inhibits tumor cell growth. This might be a good potential therapeutic target for a broad spectrum of human cancer.^[17] Therefore, the present study was carried out to draw insights regarding the relationship between the Cdc20 genes of various eukaryotes belonging to diverse taxonomic groups.^[18-20] This task was accomplished using a suite of computer software programs. The phylogenetic tree was constructed using PHYLIP 3.2 Software Package,^[21] MEGA 6.0^[22] and Mesquite 3.31^[23] software programs.

Phylogenetics is the science which studies evolutionary relationship between species. In order to make predictions about these relationships, phylogenetic trees are constructed which link the species.^[24] Maximum parsimony works by searching through possible tree structures and assigning a cost to each tree. It is based on the assumption that the mostly likely tree is the one that requires the fewest number of changes to explain the data in the alignment.^[25]

2. MATERIALS AND METHODS

2.1 Data Collection

A total of 47 nucleotide sequences of Cdc20 genes of various organisms belonging to Kingdom Protista, Fungi, Plantae and Animalia were searched in the NCBI genbank database. All the sequences were retrieved in FASTA format and were compiled into a single text file. The data set is listed in table 1.

2.2 Multiple Sequence Alignment

Multiple sequence alignment (MSA) was performed by using CLUSTAL X, which is the offline software that performs optimum alignment for multiple nucleotide and protein sequences.^[26] The required outputs for further programs were a MSA in alignment format, phylip format, Fasta format and Percent Identity Matrix. Therefore, the options in the CLUSTAL X software were adjusted as per requirements. A part of the Multiple Sequence Alignment is, shown in the figures 1(a) & (b).

2.3 Editing the Multiple sequence alignment and generating Percent Identity Matrix

The alignment obtained from the above step contained many gaps which might have caused errors in further analysis. Therefore the MSA obtained from the previous step was edited using BioEdit v7.1.9.^[27] The editing was performed by deleting the external gaps, ranging [0-1850] and [7450-8151] nucleotides, and retaining the internal gaps. The edited MSA were saved in both alignment formats viz., phylip and fasta format as per the requirements for further analysis steps.

The Percent Identity Matrix was then generated using CLUSTAL X software. The matrix is shown in figure 2.

2.4 Phylogenetic Tree Construction

2.4.1: PHYLIP Software (Version 3.69)

The multiple sequence alignment data was analyzed using three programs in PHYLIP version 3.69.^[21]

a) SEQBOOT

All the parameters were set to default except, the number of replicates being 1000. An odd number has to be selected for Random seed number and it is assigned as three. The output file viz. outfile was generated.^[28]

b) DNAPARS

The out file of SEQBOOT program acts as input for this program. Default parameters were chosen to run this program. Two files viz. outfile and outtree file were generated.^[28]

c) CONSENSE

The outtree file of DNAPARS acts as the input for this program. Two files viz. outfile and outtree file were generated. It is a Phylogenetic tree displaying the bootstrap values.^[28]

2.4.2: MEGA Software (Version 6.0)

The multiple sequence alignment data was also analyzed using MEGA software version 6.0. The edited multiple sequence alignment file in FASTA format was converted to MEGA 6.0 supported file formats. The output file viz. mas file was generated. This mas file was used to generate the phylogenetic tree by maximum parsimony method using the same MEGA 6.0 software.

The parameters used for construction of phylogenetic tree were default, except "No. of bootstrap replications" was set to 1000. With these parameters the phylogenetic tree is computed. The output file viz. Newick standard format (.nwk) was generated.

2.4.3: Mesquite software (Version 3.31)

The multiple sequence alignment data was also analyzed in Mesquite software 3.31. The edited multiple sequence alignment file in fasta format was used to build a phylogenetic tree. The following options were selected sequentially Tree value using character matrix then based on Tree length, and then Stored matrices, using NNI Rearranger which is a heuristic method to improve the likelihood of a tree and 100 equally good trees to be retained was opted.

From the total number of trees returned after the search, a consensus tree was searched. The tree generated after execution of this command depicted the relationship between the taxa on the basis of the DNA concerned.

2.5: Visualizing the Phylogenetic tree

2.5.1 Figtree Software

The resulting phylogenetic trees viz. outtree file from PHYLIP, newick standard format file from MEGA & nexus format file from MESQUITE software were visualized using Figtree v1.3.1.^[29]

3. RESULTS AND DISCUSSION

3.1 RESULTS

The group of Cdc20 genes selected from 47 eukaryotic species for the phylogenetic analysis belongs to various diverse Kingdoms of the organisms. The Multiple Sequence Alignment report [figure 1] obtained from CLUSTAL X software^[26] showed numerous gaps within the alignment (Figure. 1a). This demonstrates that the Cdc20 gene has extremely diverged during the course of its evolution. Enormous gaps pose threat while phylogenetic tree construction. Therefore, the external gapped columns were deleted using BioEdit v7.1.9 Program^[27] and the resultant alignment is represented, as in the Figure 1(b).

The edited MSA was carried forwarded to the next step of phylogenetic tree construction using PHYLIP v 3.69 software,^[21] MEGA v 6.0^[22] and Mesquite v3.31.^[23] The results of these software programs are represented in figures 3, 4, and 5. Also an attempt has been made for comparing those trees between each other.

Phylogenetic Tree constructed by PHYLIP Software

The phylogenetic tree developed by PHYLIP software is shown in figure 3; this can be divided into distinct clades P, Q, R, S. These distinct clades are divided further into sub clades and grand clades and so on.

Clade P

This clade consists of the Cdc20 sequences from diverse Kingdoms, got divided into different sub-clades as shown in the figure 3, The first sub-clade shows different branches viz. *Aptenodytes forsteri* (A19), *Heliaetus leucocephalus* (A20), *Cuculus canorus* (A21), *Gallus gallus* (A22) with constituent taxa are observed to be placed in one group, and *Gallus gallus* (A22) as out group. This might be due to the fact that the species in this clade belongs to the class Aves. The percentage of sequence divergence in this sub-clade is in between 25 to 92%.

The next sub-clade includes diverse taxa viz. *Phoenix dactylifera* (PT14), *Schizosaccharomyces pombe* (F5), *Python bivittatus* (A18) and *Salmo salar* (A24). This sub-clade has one species each from Plantae & Fungi, and two species from Animalia. The percentage range of sequence divergence in this sub clade is found to be in between 24 to 39%.

The second sub clade is observed to be the largest, consisting diverse organisms of orders Artiodactyla, Primata, Carnivora and Rodentia (Chordata: Vertebrata: Mammalia) corresponding to Kingdom Animalia. The clade consists of *Orcinus orca* (A4), *Lipotes vexillifer* (A5), *Capra hircus* (A2), *Bubalus bubalis* (A3), *Bos taurus* (A1) with *Camelus dromedarius* (A11) as an out group. All the species of this clade belong to order Artiodactyla with a 26 to 98% sequence divergence in the Cdc20 genes. A distinct bifurcation is observed in the clade containing branches *Ursus maritimus* (A9), *Homo*

sapiens (A12), *Pan troglodytes* (A13), *Macaca mulatta* (A14), *Canis lupus familiaris* (A8), *Felis catus* (A6), *Panthera pardus* (A7), with an out group branch of *Sus scrofa* (A10). The first bifurcation containing *Ursus maritimus* (A9) belonging to order Carnivora shows close relation to Primate species viz., *Homo sapiens* (A12), *Pan troglodytes* (A13), *Macaca mulatta* (A14) belongs to order Primata, with a 25 to 99% sequence divergence in the Cdc20 genes. The second bifurcated clade contains species belonging to the Order Carnivora viz. *Canis lupus familiaris* (A8), *Felis catus* (A6) and, *Panthera pardus* (A7), with 25 to 98% sequence divergence in the Cdc20 genes. *Sus scrofa* (A10) belonging to the Kingdom Animalia is an out group to this bifurcated clade. The clade containing branches with *Mesocricetus auratus* (A17), *Mus musculus* (A15) and *Rattus norvegicus* (A16) belong to Order Rodentia. *Parasteatoda tepidariorum* (A25) belongs to Order Araneae falls in the out group. The Cdc20 genes of *Chlamydomonas reinhardtii* (PT2) and *Nelumbo nucifera* (PT10) show close relationship with Kingdom Animalia.

Clade Q

The Cdc20 genes of the Kingdom Plantae are grouped in this clade. This clade with branches *Gossypium raimondii* (PT11) and *Gossypium hirsutum* (PT12), shows bifurcating taxa with *Vitis vinifera* (PT3), *Arabidopsis thaliana* (PT5) and *Solanum lycopersicum* (PT4). The Q clade has a distinct out group *Leishmania panamensis* belonging to the Kingdom Protista. The sequence divergence within the species in this clade is between 26 - 64%.

Clade R

Majority of the species grouped in this clade represented by the Cdc20 genes belong to Kingdom Fungi. The bifurcations in the clade containing branches *Saccharomyces cerevisiae* (F1), *Scheffarsomyces stipitis* (F2) and *Sugiyamaella lignohabitans* (F3), *Aspergillus fumigatus* (F4). This clade is represented by a branch *Micromonas sp.* (PT1) as the out group. An interesting features is that the Cdc20 gene of branch *Drosophila melanogaster* (A23) belonging to Kingdom Animalia [18 - 20] shows quite similarity with the Kingdom Plantae and Animalia. The sequence divergence represented within the species in this clade is between 27 - 58%.

Clade S

This clade shows exclusively plants of family Brassicaceae. The branches viz., *Brassica rapa* (PT7), *B. napus* (PT8), *B. oleracea* (PT9), followed by out groups *Erythranthe guttatus* (PT13) belongs to Kingdom Plantae and *Apis mellifera* (A26) belongs to Kingdom Animalia. The sequence divergence represented within the species in this clade is between 23 - 99%.

The phylogenetic tree shows *Hydra vulgaris* as an out group to the entire tree with branch A27.

Phylogenetic Tree constructed by MEGA Software

The phylogenetic tree developed by MEGA software is reported in figure 4. This is divided into distinct clades A, B, C, D. These distinct clades are divided further into sub-clades and grand sub-clades and so on.

Clade A

The first sub-clade is observed to be the largest one, consisting diverse organisms of order Artiodactyla, Primates, Carnivore and Rodentia corresponding to Kingdom Animalia. The clade containing branches *Bos taurus* (A1), *Capra hircus* (A2), *Bubalus bubalis* (A3), *Orcinus orca* (A4) and *Lipotes vexillifer* (A5) belong to Order Artiodactyla with a 26 to 98% sequence divergence in the Cdc20 genes.

A distinct bifurcation is observed in the clade containing branches *Felis catus* (A6), *Panthera pardus* (A7), *Canis lupus familiaris* (A8), *Ursus maritimus* (A9), with an out group branch of *Sus scrofa* (A10) and *Camelus dromedarius* (A11), having taxa with a 25 to 98% sequence divergence in the Cdc20 genes, belongs to order Carnivora and the out group branches belonging to order Artiodactyla.

The second clade contains *Homo sapiens* (A12), *Pan troglodytes* (A13) and *Macaca mulatta* (A14) having taxa belong to Order Primates, with a 25 to 99% sequence divergence in the Cdc20 genes.

The clade containing branches *Mus musculus* (A15), *Rattus norvegicus* (A16) and *Mesocricetus auratus* (A17) belong to Order Rodentia. *Python bivittatus* (A18) placed as out group branch shows close relationship with Kingdom Animalia. The next sub-clade shows different branches viz. *Aptenodytes forsteri* (A19), *Heliaetus leucocephalus* (A20), *Cuculus canorus* (A21) and *Gallus gallus* (A22) with constituent taxa placed in one group, and *Gallus gallus* (A22) as out group. This might be due to the fact that the species in this clade belongs to the Class Aves. The percentage of sequence divergence in this sub-clade is in between 26 to 92%.

Clade B

Majority species in this clade represented by the Cdc20 genes belong to Kingdoms Fungi and Plantae. The bifurcations in the clade containing branches *Saccharomyces cerevisiae* (F1), *Scheffarsomyces stipitis* (F2) and *Sugiyamaella lignohabitans* (F3), *Aspergillus fumigatus* (F4). The second sub-clade is represented by a branch in which *Micromonas sp.* (PT1), *Leishmania panamensis* (PR1), *Chlamydomonas reinhardtii* (PT2), *Drosophila melanogaster* (A23) are present. An interesting feature is that the Cdc20 gene of branch *Drosophila melanogaster* (A23) belonging to Kingdom Animalia shows quite similarity with the Kingdom Fungi and Plantae. The third sub-clade exclusively contains Kingdom Plantae, *Vitis vinifera* (PT3), *Solanum lycopersicum* (PT4), *Arabidopsis thaliana* (PT5) and *Medicago truncatula* (PT6). The percentage of sequence

divergence in this sub-clade is in between 25 to 64%. *Salmo salar* (A24) belonging to Kingdom Animalia is an out group branch.

Clade C

This clade exclusively contains branches of Kingdom Plantae viz. *Brassica rapa* (PT7), *B. napus* (PT8), *B. oleracea* (PT9) and *Nelumbo nucifera* (PT10), followed by next sub-clade with branches *Gossypium raimondii* (PT11), *G. hirsutum* (PT12) and *Erythranthe guttatus* (PT13). *Parasteatoda tepidariorum* (A25) belonging to Order Araneae falls in the out group along with *Phoenix dactylifera* (PT14).

Clade D

This clade contains branches with *Apis mellifera* (A26) and *Hydra vulgaris* (A27) both belonging to Kingdom Animalia. *Schizosaccharomyces pombe* (F5) as an out group branch to the total phylogenetic tree.

Phylogenetic Tree constructed by MESQUITE Software

The phylogenetic tree developed by MESQUITE software is depicted in figure 5. This can be divided into four distinct clades viz. K, L, M and N. These clades are divided further into sub-clades, grand clades and so on.

Clade K

The first sub-clade is the largest sub-clade, consisting of diverse organisms belonging to the Orders Artiodactyla, Primata, Carnivora and Rodentia of Kingdom Animalia. The clade containing branches *Bos taurus* (A1), *Bubalus bubalis* (A3), *Capra hircus* (A2), *Lipotes vexillifer* (A5), *Orcinus orca* (A4), *Camelus dromedarius* (A11), *Sus scrofa* (A10) belong to order Artiodactyla with 26 to 98% sequence divergence in the Cdc20 genes. The second sub-clade has two bifurcations. First bifurcation containing branches *Mus musculus* (A15), *Rattus norvegicus* (A16) and *Mesocricetus auratus* (A17) belong to Order Rodentia (Mammalia) and the second one with branches *Pan troglodytes* (A13), *Homo sapiens* (A12) and *Macaca mulatta* (A14) having taxa belonging to Order Primata (Mammalia), with a 25 to 99% sequence divergence in the Cdc20 genes. The next sub-clade *Panthera pardus* (A7), *Felis catus* (A6), with *Canis lupus familiaris* (A8) and *Ursus maritimus* (A9) as out group branches, having taxa with a 25 to 98% sequence divergence in the Cdc20 genes, belongs to Order Carnivora.

The next sub-clade, with constituent taxa *Heliaetus leucocephalus* (A20), *Aptenodytes forsteri* (A19) and *Cuculus canorus* (A21) are placed in one group. *Gallus gallus* (A22) and *Cuculus canorus* (A21) emerged as out group. This might be due to the fact that the species in this sub-clade belong to the Class Aves. The percentage of sequence divergence in this sub-clade is in between 26 to 92%. *Chlamydomonas reinhardtii* (PT2) belonging to Kingdom Plantae shows unusual relation with Kingdom Animalia and is placed as an out group branch with

reference to divergence of Cdc20 gene. *Python bivittatus* (A18) falls in the out group along with *Salmo salar* (A24), *Hydra vulgaris* (A27) and *Parasteatoda tepidariorum* (A25) belonging to Order Araneae.

Clade L

The second clade has many sub-clades which further have grand clades. The first sub-clade is divided into two grand clades with first one having *Saccharomyces cerevisiae* (F1), *Scheffarsomyces stipitis* (F2) and second having *Sugiyamaella lignohabitans* (F3), *Aspergillus fumigatus* (F4). The second sub-clade is represented by a branch *Leishmania panamensis* (PR1) belonging to Kingdom Protista and *Micromonas* sp. (PT1), this shows that Cdc20 gene has close relation between protists and early plants. *Drosophila melanogaster* (A23) placed as out group branch in this sub-clade. Next comes sub-clade with branches *Solanum lycopersicum* (PT4), *Vitis vinifera* (PT3) and *Arabidopsis thaliana* (PT5), with *Medicago truncatula* (PT6) as out group.

Clade M

Apis mellifera (A26) and *Schizosaccharomyces pombe* (F5) are grouped into one sub-clade. This is followed by next sub-clade with branches *Gossypium hirsutum* (PT12), *G. raimondii* (PT11) and *Erythranthe guttatus* (PT13). *Phoenix dactylifera* (PT14) and *Nelumbo nucifera* (PT10), are placed as out group branching.

Clade N

B. rapa (PT7), *B. napus* (PT8) are grouped into one clade. The phylogenetic tree shows *B. oleracea* (PT9) as an out group to the entire tree with single branch, all the three belonging to Kingdom Plantae and Brassicaceae family.

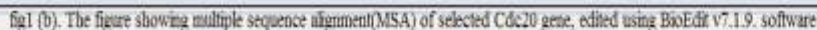
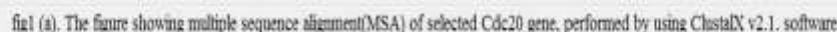
Comparative account on phylogenetic trees developed using PHYLIP, MEGA and MESQUITE software

The sequences of Cdc20 gene showed some interesting results after building phylogenetic trees. The comparative analysis of these three phylogenetic trees derived from the three different software programs, the sequences of Cdc20 gene of Kingdom Animalia segregated distinctly over the dendrogram into a clade. Moreover, even within this clade, the sequences clustered according to their phylogenetic relations such as sequences of organisms belonging to same order clustered together. In all the three trees Orders Artiodactyla, Rodentia, Aves and Primata the clustering of branches is similar, although the positions of the branches are vary in these trees,. There is an exception in the Order Artiodactyla (Mammalia), that *Sus scrofa* (A10), *Camelus dromedarius* (A11) are placed as out group branches in all three trees. The Cdc20 sequences of the Order Carnivora (Mammalia) shows less distance in the arrangement but the arrangement of the branches varies in the trees. The most common similarity among the trees is that the Cdc20 sequences of *Leishmania panamensis* (PR1) is placed nearer to Kingdom Plantae, suggesting that early plants might have evolved from protists justifying RH Whittaker's five Kingdoms classification. Even though *Drosophila melanogaster* (A23), *Parasteatoda tepidariorum* (A25) and *Apis mellifera* (A26) belong to Kingdom Animalia and Class Insecta, were branched separately in between the branches of Kingdom Plantae in all the three trees. In the branching of Kingdom Fungi, *Saccharomyces cerevisiae* (F1), *Scheffarsomyces stipitis* (F2) and *Sugiyamaella lignohabitans* (F3), *Aspergillus fumigatus* (F4) are placed in a single sub-clade but only *Schizosaccharomyces pombe* (F5) is placed separately, a common factor among the trees derived adopting the three different softwares. All the three phylogenetic trees developed from different software programs shows Kingdom Animalia latest to be evolved during the course of evolution of Cdc20 gene supporting the view of RH Whittaker's classification.^[19]

Table 1: The table shows the Accession numbers of Cdc20 genes for 47 species retrieved from Genbank nucleotide sequences database.

S.No	Kingdom	Scientific name	Accession number	Abbreviation	Branch name
1.	Protista	Leishmania panamensis	NC_025870.1	Leishmania	PR1
2.	Fungi	Saccharomyces cerevisiae	NC_001139.9	Saccharomy	F1
3.	Fungi	Scheffarsomyces stipitis	NC_009045.1	Scheffarso	F2
4.	Fungi	Sugiyamaella lignohabitans	NC_031674.1	Sugiyamael	F3
5.	Fungi	Aspergillus fumigatus	NC_007194.1	Aspergillu	F4
6.	Fungi	Schizosaccharomycespombe	NC_003423.3	Schizosacc	F5
7.	Plantae	Micromonas sp.	NC_013042.1	Micromonas	PT1
8.	Plantae	Chlamydomonas reinhardtii	NW_001843697.1	Chlamydomo	PT2
9.	Plantae	Vitis vinifera	NC_010364530	Vitis_vini	PT3
10.	Plantae	Solanum lycopersicum	NC_015440.2	Solanum_ly	PT4
11.	Plantae	Arabidopsis thaliana	NC_003075.7	Arabidopsi	PT5
12.	Plantae	Medicago truncatula	NC_016413.2	Medicago_t	PT6
13.	Plantae	Brassica rapa	NC_024802.1	Brassica_r	PT7
14.	Plantae	Brassica napus	NC_027771.1	Brassica_n	PT8

15.	Plantae	Brassica oleracea	NC_027755.1	Brassica_o	PT9
16.	Plantae	Nelumbo nucifera	NW_010729097.1	Nelumbo_nu	PT10
17.	Plantae	Gossypium raimondii	NC_026936.1	Gossypium_r	PT11
18.	Plantae	Gossypium hirsutum	NW_015962259.1	Gossypium_h	PT12
19.	Plantae	Erythranthe guttatus	NW_01219497.1	Erythranth	PT13
20.	Plantae	Phoenix dactylifera	NW_008246865.1	Phoenix_da	PT14
21.	Animalia	Bos taurus	AC_000160.1	Bos_taurus	A1
22.	Animalia	Capra hircus	NC_030810.1	Capra_hirc	A2
23.	Animalia	Bubalus bubalis	NW_005785311.1	Bubalus_bu	A3
24.	Animalia	Orcinus orca	NW_004438433.1	Orcinus_or	A4
25.	Animalia	Lipotes vexillifer	NW_006796183.1	Lipotes_ve	A5
26.	Animalia	Felis catus	NC_018730.2	Felis_catu	A6
27.	Animalia	Panthera pardus	NW_017619967.1	Panthera_p	A7
28.	Animalia	Canis lupus familiaris	NC_006597.3	Canis_lupu	A8
29.	Animalia	Ursus maritimus	NW_007907174.1	Ursus_mari	A9
30.	Animalia	Sus scrofa	NC_010448.3	Sus_scrofa	A10
31.	Animalia	Camelus dromedarius	NW_011591250.1	Camelus_dr	A11
32.	Animalia	Homo sapiens	NC_000001.11	Homo_sapie	A12
33.	Animalia	Pan troglodytes	NC_006468.4	Pan_troglo	A13
34.	Animalia	Macaca mulatta	NC_027893.1	Macaca_mul	A14
35.	Animalia	Mus musculus	NC_000070.6	Mus_muscul	A15
36.	Animalia	Rattus norvegicus	NC_005104.4	Rattus_nor	A16
37.	Animalia	Mesocricetus auratus	NW_004801760.1	Mesocricet	A17
38.	Animalia	Python bivittatus	NW_006539021.1	Python_biv	A18
39.	Animalia	Aptenodytes forsteri	NW_008796165.1	Aptenodyte	A19
40.	Animalia	Heliaeetus leucocephalus	NW_010973284.1	Heliaeetus	A20
41.	Animalia	Cuculus canorus	NW_009244581.1	Cuculus_ca	A21
42.	Animalia	Gallus gallus	NC_006095.4	Gallus_gal	A22
43.	Animalia	Drosophila melanogaster	NT_033779.5	Drosophila	A23
44.	Animalia	Salmo salar	NC_027302.1	Salmo_sala	A24
45.	Animalia	Parasteatoda tepidariorum	NW_015470508.1	Parasteato	A25
46.	Animalia	Apis mellifera	NC_007074.3	Apis_melli	A26
47.	Animalia	Hydra vulgaris	NW_004169671.1	Hydra_vulg	A27



Percent Identity Matrix by Clustal 2.1

[illegible]

fig2. The figure showing percentage identity matrix generated using ClustalX v2.1

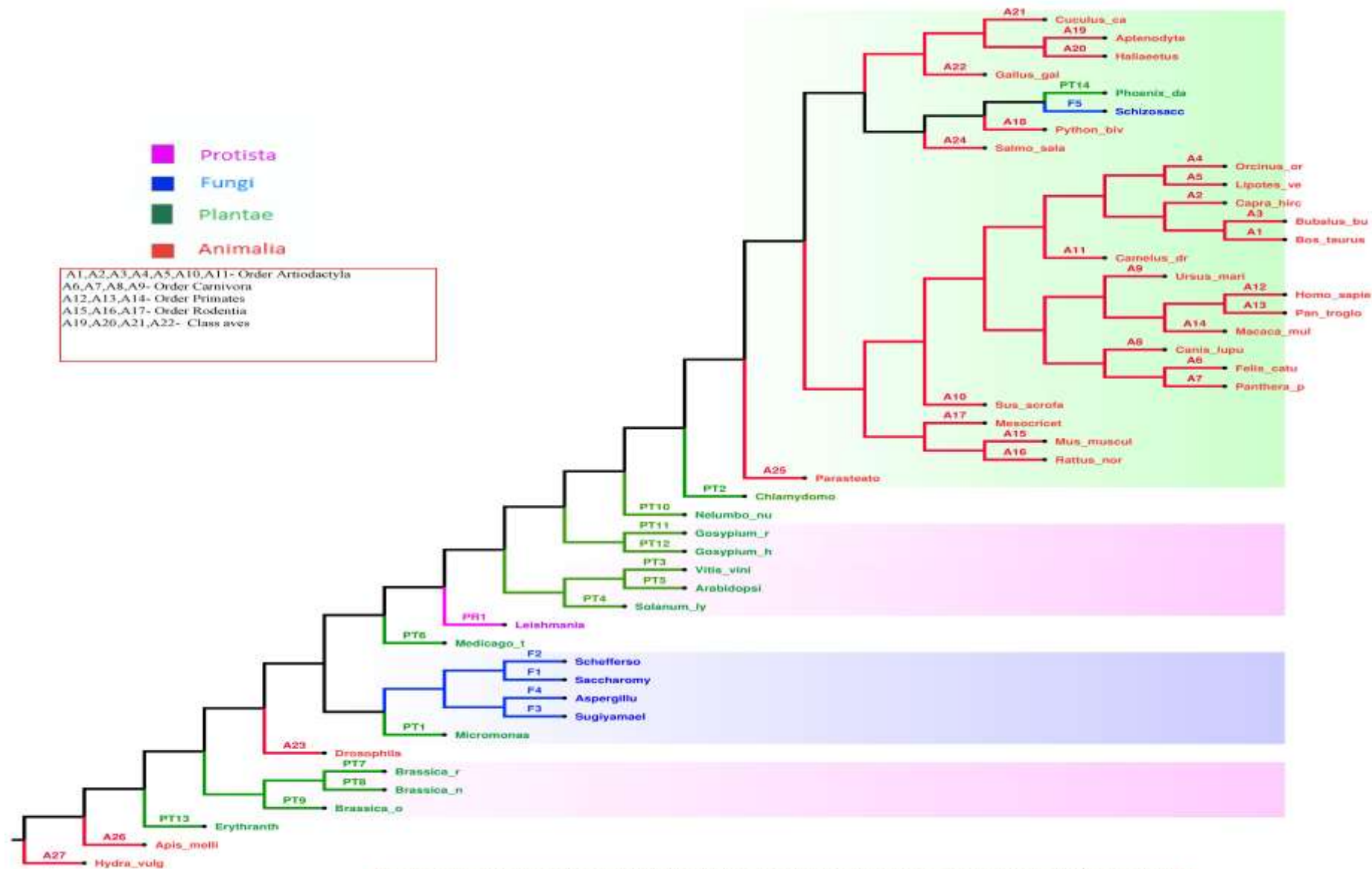


fig3. The figure showing the phylogenetic tree representing the divergence of the gene Cdc20, generated using PHYLIP v3.69 software

3.0



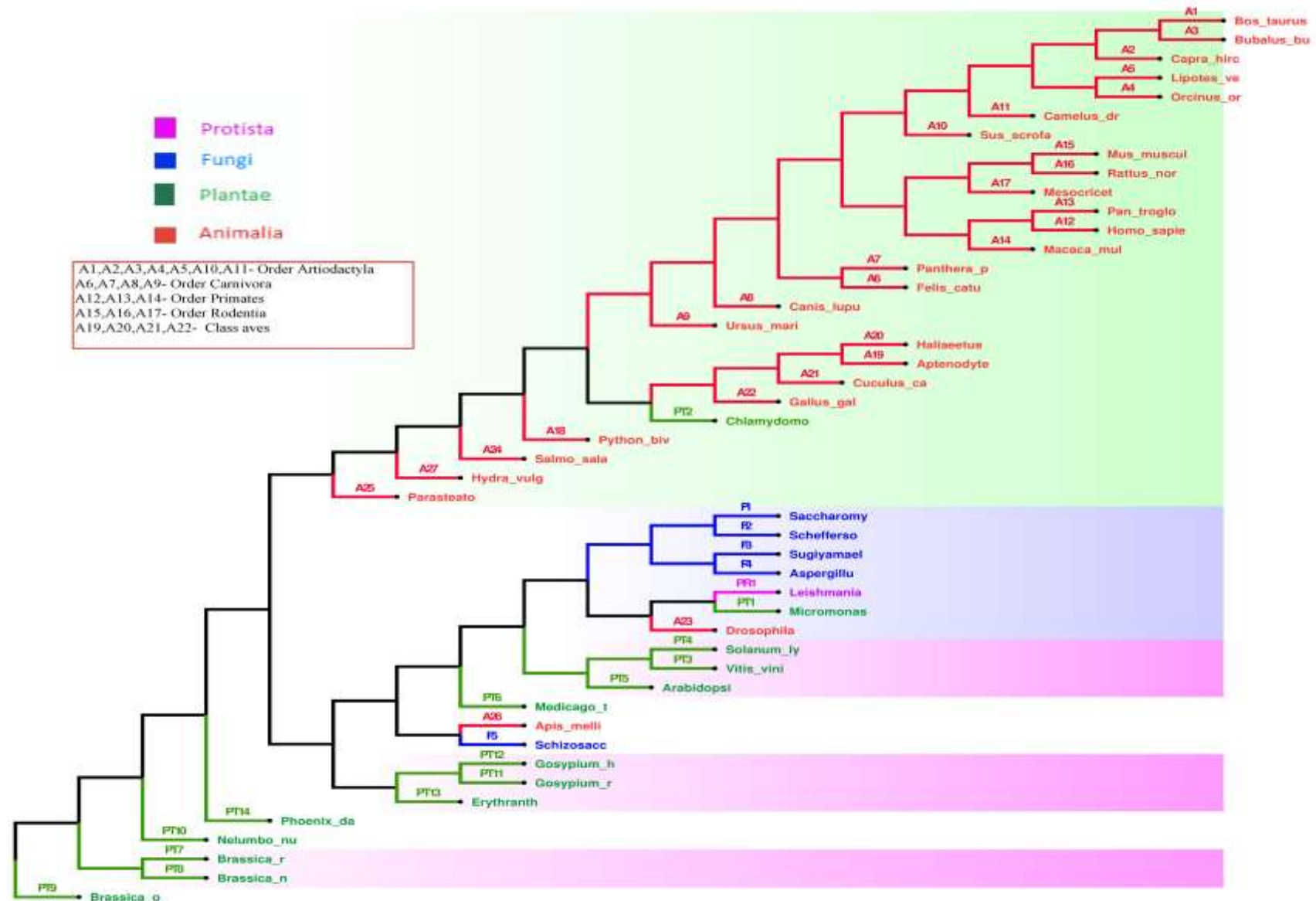


fig5. The figure showing the phylogenetic tree representing the divergence of the gene Cdc20, generated using Mesquite v3.31

2.0

3.2 DISCUSSION

The study was carried out using certain computer software programs. The comparative analysis of Cdc20 gene in various Eukaryotes using three software programs, i.e. PHYLIP, MEGA and MESQUITE showed quite interesting and unexpected results. Although the output of the three programs is variable, similar results were obtained for majority of species. The constructed phylogenetic trees showed *Drosophila melanogaster* (A23), *Parasteatoda tepidariorum* (A25) and *Apis mellifera* (A26) belonging to Kingdom Animalia and class Insecta were branched separately in between the branches of Kingdom Plantae in the three trees obtained through the three softwares. This seems peculiar in insects having closer relation with Kingdom Plantae with reference to divergence of Cdc20 gene. In the same way, there are such certain peculiar patterns of branching of the Cdc20 gene in the resulting phylogenetic trees.

Based on the literature survey, the gene Cdc20 plays a critical oncogenic role in the development and progression of human cancers, which when inhibited by certain inhibitors, can play as a novel therapeutic target for cancer.^[16] Such results help the researchers to draw insights for gene therapy and the conserved regions of Cdc20 gene of lower eukaryotes can be helpful for understanding and implementing required changes for the development of therapeutic ability of Cdc20 gene and its inhibitors. In PhyloOncology, from grouping subsets of cancer samples to tracing subclonal evolution during cancer progression and metastasis, the use of phylogenetics is a powerful systems biology approach. Well-developed phylogenetic applications provide fast, robust approaches to analyze high-dimensional, heterogeneous cancer data sets.^[30]

4. CONCLUSION

In this study, evolutionary relationship of Cdc20 among 47 eukaryotic species enunciated the importance of identifying various unique features such as variation and conserved regions. The Phylogenetic trees showed that Cdc20 gene being 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 such as genus, family, order or class etc. Some clusters showed up to 98-99% sequence identity in organisms belonging to similar order. 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 for exploring insights among them to initiate research work by comparative approach. Based on literature review, the gene Cdc20 plays a critical oncogenic role in the development and progression of human cancers. With the help of PhyloOncology, an emerging subject in phylogenetics, further analysis and in depth investigation can be performed for evolutionary analysis of cancer among the eukaryotes.

5. ACKNOWLEDGEMENTS

R Prasuja and M Jagannath are thankful to the Department of Biotechnology, Bioinformatics Division, Govt. of India, New Delhi for financial support and also thankful to Ms. K Umadevi, Associate Professor, Department of Marine Living Resources & Coordinator DBT-BIF, Andhra University, and Dr. Rupesh S Badere, Assistant Professor, Department of Botany, RTM Nagpur University, Nagpur, for final correction of manuscript.

6. REFERENCES

1. Database for Cdc20 gene sequences, compiled by National Center for Biotechnology Information (NCBI), <https://www.ncbi.nlm.nih.gov/gene/?term=CDC20>.
2. Basic information related to Cdc20 gene, compiled by Hugo Gene Nomenclature Committee(HGNC),https://www.genenames.org/cgi-bin/gene_symbol_report?hgnc_id=1723.
3. Leland H. Hartwell. Joseph Culotti, Brian Reid, Genetic Control of the Cell-Division Cycle in Yeast, I. Detection of Mutants. Proceedings of the National Academy of Sciences., 1970; 66(2): 352-359.
4. Leland H. Hartwell, Robert K. Mortimer, Joseph Culotti and Marilyn Culotti. Genetic control of the cell-division cycle in the yeast: v.genetic analysis of Cdc mutants. Genetics., 1973; 74: 267-286.
5. Jasminder Weinstein, Frederick W. Jacobsen, Jennifer Hsu-chen, TerryWu, Linda G. Baum. A Novel Mammalian Protein, p55CDC, Present in dividing cells is associated with protein kinase activity and has homology to the *Saccharomyces cerevisiae* cell division cycle proteins Cdc2O and Cdc4, Molecular and Cellular Biology., 1994; 3350-3363.
6. Jasminder Weinstein. Cell Cycle-regulated Expression, Phosphorylation, and Degradation of p55Cdc- A Mammalian Homolog of CDC20/Fizzy/slp1. The Journal of Biological Chemistry., 1997; 272(45): 28501-28511.
7. Jing Pan and Rey-Huei Chen. Spindle checkpoint regulates Cdc20p stability in *Saccharomyces cerevisiae*. Genes and Development., 2004; 18:1439-1451.
8. Marko J. Kallio, Victoria A. Beardmore, Jasminder Weinstein, Gary J. Gorbsky. Rapid microtubule-independent dynamics of Cdc20 at kinetochores and centrosomes in mammalian cells. The Journal of Cell Biology, 2002; 158(5) 841-847.
9. Hong Hwa Lim, Phuay-Yee Goh, Uttam surana. Cdc20 is essential for the cyclosome-mediated proteolysis of both Pds1 and Clb2 during M phase in budding yeast. Current Biology., 1998; 8: 231-23.
10. Randall W. King, Jan-Michael Peters, Stuart Tugendreich, Mark Rolfe, Philip Hieter, Marc W.Kirschner. A 20s Complex containing Cdc27 and Cdc16 catalyzes the mitosis-specific conjugation of

- ubiquitin to cyclin B. *Cell*. 1995; Vol. 81, 2: 279-288.
11. Valery Sudakin, Dvorah Ganoth, Aviva Dahan, Hannah Heller, Judith Hershko, Francis C. Luca, Joan V. Ruderman, Avram Hershko. The Cyclosome, a Large Complex Containing Cyclin-Selective Ubiquitin Ligase Activity, Targets Cyclins for Destruction at the End of Mitosis. *Molecular Biology of the Cell*, 1995; 6: 185-198.
 12. Leifu Chang, Ziguo Zhang, Jing Yang, Stephen H. McLaughlin, David Barford. Molecular architecture and mechanism of the anaphase-promoting complex. *Nature*, 2014; 513: 388-393.
 13. Guowei Fang, Hong tao Yu, Marc W. Kirschner. The checkpoint protein MAD2 and the mitotic regulator CDC20 form a ternary complex with the anaphase-promoting complex to control anaphase initiation. *Genes and Development*, 1998; 12: 1871-1883.
 14. Katja Wassmann, Robert Benezra. Mad2 transiently associates with an APC^{Cyp55Cdc} complex during mitosis. *Cell Biology*, 1998; 95: 11193-11198.
 15. Yongke Zhang, Emma Lees. Identification of an overlapping binding domain on Cdc20 for Mad2 and Anaphase-Promoting complex: Model for spindle checkpoint regulation. *Molecular Cell Biology*, 2001; 21(15): 5190-5199.
 16. Zhiwei Wang, Lixin Wan, Jiateng Zhong, Hiroyuki Inuzuka, Pengda Liu, Fazlul H. Sarkar, Wenyi Wei. Cdc20: a potential novel therapeutic target for cancer treatment. *Current Pharma Design*, 2013; 19(18): 3210-3214.
 17. T Kidokoro, C Tanikawa, Y Furukawa, T Katagiri, Y Nakamura, K Matsuda. CDC20, a potential cancer therapeutic target, is negatively regulated by p53. *Oncogene*, 2008; 27: 1562-1571.
 18. R. H. Whittaker. On the broad classification of organisms. *Quart. Rev. Biology*, 1959; 34: 210-226.
 19. R. H. Whittaker. New concepts of Kingdoms of organisms. *Science*, 1969; 163: 150-160.
 20. L. Margulis, K. V. Schwartz, Diana L. Lipscomb. Five Kingdoms: An Illustrated Guide to the Phyla of Life on Earth. *Systematic Zoology*. (Dec., 1982), 1982; 31(4): 523-526.
 21. Joseph Felsenstein. PHYLIP - Phylogeny Inference Package (Version 3.69). *Cladistics*. 1989; Vol. 5, pp. 164-166.
 22. Koichiro Tamura, Glen Stecher, Daniel Peterson, Alan Filipski, and Sudhir Kumar. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*. 2013; 30: 2725-2729.
 23. Maddison, W. P. and D.R. Maddison. Mesquite: a modular system for evolutionary analysis. 2017; Version 3.31 <http://mesquiteproject.org>
 24. Jeffrey Rizzo, Eric C. Rouchka. Review of Phylogenetic Tree Construction. University of Louisville Bioinformatics Laboratory Technical Report Series, 2007 Report number TR-ULBL-2007-01
 25. B. G. Hall, *Phylogenetic Trees Made Easy: A How-to Manual*, 3 Edition, Sinauer Associates, Inc., 2007.
 26. M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins. Clustal W and Clustal X version 2.0. *Bioinformatics*, 1 November 2007; 23(21): 2947-2948, <https://doi.org/10.1093/bioinformatics/btm404>
 27. Hall TA. "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT". *Nucl. Acids. Symp. Ser.*, 1999; 41: 95-98.
 28. M. Jagannath, K. Umadevi, Subhalaxmi Nayak. Phylogenetic analysis of mitochondrial 16s rRNA and NADH Dehydrogenase genes of Takifugu (Pisces: *Tetraodontiformes: Tetraodontidae*). *Online Journal of Bioinformatics*, 2011; 12(2): 255-273.
 29. Rambaut A. Institute of Evolutionary Biology. University of Edinburgh. Edinburgh. <http://tree.bio.ed.ac.uk/software/figtree/>.
 30. Jason A. Somarelli, Kathryn E. Ware, Rumen Kostadinov, Jeffrey M. Robinson, Hakima Amri, Mones Abu-Asab Nicolaas Fourie, Rui Diogo, David Swofford, Jeffrey P. Townsend. *PhyloOncology: Understanding cancer through phylogenetic analysis*. *Biochimica et Biophysica Acta*, 2016; <http://dx.doi.org/10.1016/j.bbcan.2016.10.006>.