International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-2, Issue-6, November 2014

Phylogenetic Analysis of E6/E7 (Homo sapiens) Protein of Human Papillomavirus (HPV) Through Bioinformatics

Ashok Kumar Saxena, Vijay Laxmi Saxena, Bhartesh Kumar

Abstract— **Human papillomaviruses (HPVs) are a heterogeneous group of small dsDNA viruses which cause a variety of proliferative epithelial lesions at specific anatomical sites. In order to classify HPV protein types, phylogenetic trees were constructed based on amino acid sequence alignments using parsimony and distance matrix algorithms. Simultaneous phylogenetic analysis of E6 and E7 proteins of 42 different HPV strains which mainly found in Homo sapiens was carried out in detail. Both E6 and E7 proteins of different HPV strains showed evolutionary divergence into two major distinct lineages. While E6 protein was further differentiated into 7 smaller lineages, E7 differentiated into 8 lineages. Multiple Sequence Alignment (MSA) results revealed their amino acid profiles demonstrated conserved lineage-specific substitutions independently. The resulting phylogenetic trees provide a classification of the HPVs into specific groups encompassing the known tissue tropism and oncogenic potential of each HPV type. The implications of a phylogenetic taxonomy on the diagnostic detection of HPVs and the concept of different HPV species are discussed. Dendrogram topologies of E6 and E7 proteins among different HPV were very similar which showed that in most of the HPV, both the proteins were evolved in a similar manner. Also, similar phylogenetic profiles among different HPV types having fully / highly conserved residues were observed, suggesting possible functional similarities.**

Index Terms— **E6 & E7 protein, Human Papillomavirus, MSA, Phylogenetic analysis.**

I. INTRODUCTION

Phylogenetics is the study of evolutionary relationships. Phylogenetic analysis is the means of inferring or estimating these relationships. The evolutionary history inferred from phylogenetic analysis is usually depicted as branching, treelike diagrams that represent an estimated pedigree of the inherited relationships among molecules (‘‘protein trees’’), organisms, or both. Phylogenetics is sometimes called cladistics because the word ‘‘clade,’’ a set of descendants from a single ancestor, is derived from the Greek word for branch. However, cladistics is a particular method of hypothesizing about evolutionary relationships [1] [2].

Human papillomavirus (HPV) is a virus from the papillomavirus family that is capable of infecting humans. Like all papillomaviruses, HPVs establish productive infections only in keratinocytes of the skin or mucous membranes. While the majority of the known types of HPV cause no symptoms in most people, some types can cause warts (verrucae), while others can – in a minority of cases – lead to cancers of the cervix, vulva, vagina, penis, oropharynx and anus[3]. Recently, HPV has been linked with an increased risk of cardiovascular disease [4]. In addition, HPV 16 and 18 infections are strongly associated with an increased odds ratio of developing oropharyngeal (throat) cancer [5]. More than 30 to 40 types of HPV are typically transmitted through sexual contact and infect the anogenital region. Some sexually transmitted HPV types may cause genital warts. Persistent infection with "high-risk" HPV types — different from the ones that cause skin warts — may progress to precancerous lesions and invasive cancer [6]. HPV infection is a cause of nearly all cases of cervical cancer [7]. However, most infections with these types do not cause disease. Most HPV infections in young females are temporary and have little long-term significance. Seventy percent of infections are gone in 1 year and ninety percent in 2 years [8]. However, when the infection persists — in 5% to 10% of infected women — there is high risk of developing precancerous lesions of the cervix, which can progress to invasive cervical cancer. This process usually takes 10–15 years, providing many opportunities for detection and treatment of the pre-cancerous lesion.

The two primary oncoproteins of high risk HPV types are E6 and E7. The “E” designation indicates that these two proteins are expressed early in the HPV life cycle, while the "L" designation indicates late expression [9]. The HPV genome is composed of six early (E1, E2, E4, E5, E6, and E7) ORFs, two late (L1 and L2) ORFs, and a non-coding long control region (LCR) [10]. After the host cell is infected viral early promoter is activated and a polycistronic

Manuscript received November 25, 2014.

Ashok Kumar Saxena, Department of Zoology, D.A.V. College civil Line Kanpur (U.P), India, 05122530657, (e-mail: vijayashokkanpur@gmail.com).

Vijay Laxmi Saxena, Coordinator: Bioinforamtics Infrastructure Facility, Centre of DBT (Govt.Iindia), Department of Zoology, D.G.P.G. College civil Line Kanpur (U.P), India, 05122530657, (e-mail: vijayashokkanpur@gmail.com).

Bhartesh Kumar, Bioinforamtics Infrastructure Facility, Centre of DBT (Govt.Iindia), D.G.P.G.College civil Line Kanpur (U.P), India, 9058308030, (e-mail: bhartesh.mbi@gmail.com).

primary RNA containing all six early ORFs is transcribed. This polycistronic RNA then undergoes active RNA splicing to generate multiple isoforms of mRNAs [11]. One of the spliced isoform RNAs, E6*I, serves as an E7 mRNA to translate E7 protein [12]. However, viral early transcription subjects to viral E2 regulation and high E2 levels repress the transcription. E6 and E7 expression promote to cellular proliferation and the chance of malignancy. The degree to which E6 and E7 are expressed is correlated with the type of cervical lesion that can ultimately develop [13]. The E6/E7 proteins inactivate two tumor suppressor proteins, p53 (inactivated by E6) and pRb (inactivated by E7) [14]. E6 in association with host E6-associated protein, which has ubiquitin ligase activity, acts to ubiquitinate p53, leading to its proteosomal degradation. E7 (in oncogenic HPVs) acts as the primary transforming protein. E7 competes for retinoblastoma protein (pRb) binding, freeing the transcription factor E2F to transactivate its targets, thus pushing the cell cycle forward. In the upper layers of the host epithelium, the late genes L1 and L2 are transcribed/translated and serve as structural proteins that encapsidate the amplified viral genomes. Once the genome is encapsidated, the capsid appears to undergo a redox-dependent assembly/maturation event, which is tied to a natural redox gradient that spans both suprabasal and cornified epithelial tissue layers. This assembly/maturation event stabilizes virions, and increases their specific infectivity [15]. Virions can then be sloughed off in the dead squamous of the host epithelium and the viral lifecycle continues [16]. A 2010 study has found that E6 and E7 are involved in beta-catenin nuclear accumulation and activation of Wnt signaling in HPV-induced cancers [17].

II. MATERIAL

A. Molecular Evolution

Evolution is a process by which the traits of a population change from one generation to another. In On the Origin of Species by Means of Natural Selection, Darwin proposed that, given overwhelming evidence from his extensive comparative analysis of living specimens and fossils, all living organisms descended from a common ancestor. Illustration is a tree-like structure that suggests how slow and successive modifications could lead to the extreme variations seen in species today [18] [19]. Darwin's theory of evolution is based on three underlying principles: variation in traits exist among individuals within a population, these variations can be passed from one generation to the next via inheritance, and that some forms of inherited traits provide individuals a higher chance of survival and reproduction than others [19].

B. WHAT IS PHYLOGENY ?

According to modern evolutionary theory, all organisms on earth have descended from a common ancestor, which means that any set of species, extant or extinct, is related. This relationship is called a phylogeny, and is represented by phylogenetic trees, which graphically represent the evolutionary history related to the species of interest. Phylogenetics infers trees from observations about existing organisms using morphological, physiological, and molecular characteristics. The "tree of life" represents a phylogeny of all organisms, living and extinct. Other, more specialized species and molecular phylogenies are used to support comparative studies, test biogeographic hypotheses, evaluate mode and timing of speciation, infer amino acid sequence of extinct proteins, track the evolution of diseases, and even provide evidence in criminal cases [20].

C. Methods of Phylogenetic Analysis

Although the nature and scope of phylogenetic studies may vary significantly and require different datasets and computational methods, the basic steps in any phylogenetic analysis remain the same: assemble and align a dataset, build (estimate) phylogenetic trees from sequences using computational methods and stochastic models, and statistically test and assess the estimated trees [21]- [23].

a) Building Phylogenetic Trees

To build phylogenetic trees, statistical methods are applied to determine the tree topology and calculate the branch lengths that best describe the phylogenetic relationships of the aligned sequences in a dataset. Many different methods for building trees exist and no single method performs well for all types of trees and datasets. The most common computational methods applied include distance-matrix methods, and discrete data methods, such as maximum parsimony and maximum likehood [24].

D. What is Phylogenetic tree?

A phylogenetic tree is a statement about the evolutionary relationship between a set of homologous characters of one or several organisms. Homology according to Fitch is the relationship of two characters that have descended usually with divergence, from a common ancestral character. The characters can be any genic (gene sequence, protein sequence), structural (i.e. morphological) or behavioural feature of an organism.

a) Maximum parsimony phylogenetics

Maximum parsimony algorithms search for the minimum number of genetic events (nucleotide substitutions or amino acid changes) to infer the most parsimonious tree from a set of sequences. The phylogenetic trees were generated with the software package PAUP 3.0 (Phylogenetic Analysis

Using Parsimony), using the heuristic search option with random stepwise addition of sequences (100 replications, holding five trees at every step), and 'nearest neighbor interchange' branch swapping. Two other branch swapping or rearrangement algorithms ('subtree pruning-regrafting' and 'tree bisection-reconnec- tion') failed to identify more parsimonious trees [25].

b) Distance matrix analysis

Distance matrix methods convert sequence data into a set of discrete pairwise distance values. Each pairwise similarity score was computed as the number of exactly matching characters in an optimal alignment, minus a fixed penalty for every gap, using the heuristic algorithm [26]. The topology of the phylogenetic tree was resolved by clustering the most similar sequences according to the unweighted pair group maximum average (UPGMA) analysis [27][28]. Horizontal branch lengths in UPGMA-derived trees are not to scale and therefore do not metrically represent evolutionary change.

III. HUMAN PAPILLOMA VIRUS (HPV)

Human papillomaviruses (HPVs) are a heterogeneous group of small dsDNA viruses which cause a variety of proliferative epithelial lesions at specific anatomical sites. Human Papillomavirus (HPV) is one of the most common virus groups in the world today affecting the skin and mucosal areas of the body. It was assumed until recently that all infectious wart-like lesions in humans were caused by a single papillomavirus infecting different anatomical sites **[29]**. The biological manifestations attributed to HPVs are manifold and are of interest to an unusually wide range of clinicians and laboratory scientists. HPVs can cause benign lesions such as common, fiat or plantar warts, genital HPV infection, including condyloma acuminatum, is now recognized as the most common sexually transmitted disease, laryngeal papilloma's are often recurrent and can become sufficiently large to cause airway obstruction and death, oral focal epithelial hyperplasia, oral condylomata, inverted nasal papillomas and conjunctival papillomas have also been shown to be HPV-related. Much of the current interest in papillomavirus research can be attributed to the recent association of HPVs with premalignant and malignant epithelial lesions.

A. E6 Protein

In the HPV (Human Papillomavirus) total 43 sequence of E6 protein of Homo sapiens are obtained. E6 is a 151 amino-acid peptide that incorporates a type 1 motif with a consensus sequence – (T/S)-(X)-(V/I) - COOH. It also has two zinc finger motifs. E6 is of particular interest because it appears to have multiple roles in the cell and to interact with many other proteins. Its major role, however, is to mediate the degradation of p53, a major tumor suppressor protein, reducing the cell's ability to respond to DNA damage. E6 has also been shown to target other cellular proteins, thereby altering several metabolic pathways. One such target is NFX1-91, which normally represses production of telomerase, a protein that allows cells to divide an unlimited number of times. When NFX1-91 is degraded by E6, telomerase levels increase, inactivating a major mechanism keeping cell growth in check. Additionally, E6 can act as a transcriptional cofactor—specifically, a transcription activator—when interacting with the cellular transcription factor, E2F1/DP1. E6 can also bind to PDZ-domains, short sequences which are often found in signaling proteins. E6 proteins also interact with the MAGUK (membrane-associated guanylate kinase family) proteins. These proteins, including MAGI-1, MAGI-2, and MAGI-3 are usually structural proteins, and can help with signaling. When E6 complexes with the PDZ domains on the MAGI proteins, it distorts their shape and thereby impedes their function. Overall, the E6 protein serves to impede normal protein activity in such a way as to allow a cell to grow and multiply at the increased rate characteristic of cancer. Since the expression of E6 is strictly required for maintenance of a malignant phenotype in HPV-induced cancers, it is an appealing target of therapeutic HPV vaccines designed to eradicate established cervical cancer tumors.

B)E7 Protein

In the HPV (Human Papillomavirus) total 18 sequences of E7 protein of Homo sapiens have been obtained. In most papillomavirus types, the primary function of the E7 protein is to inactivate members of the pRb family of tumor suppressor proteins. Together with E6, E7 serves to prevent cell death (apoptosis*)* and promote cell cycle progression, thus priming the cell for replication of the viral DNA. E7 also participates in immortalization of infected cells by activating cellular telomerase. Like E6, E7 is the subject of intense research interest and is believed to exert a wide variety of other effects on infected cells. As with E6, the ongoing expression of E7 is required for survival of cancer cell lines, such as **HeLa**, that are derived from HPV-induced tumors.

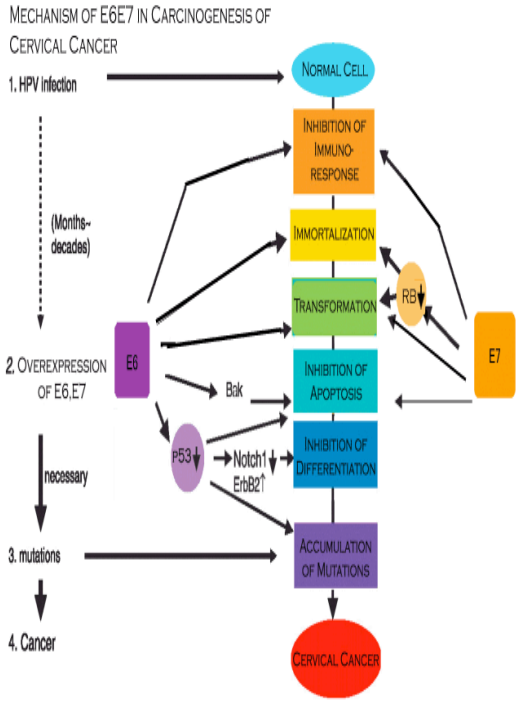


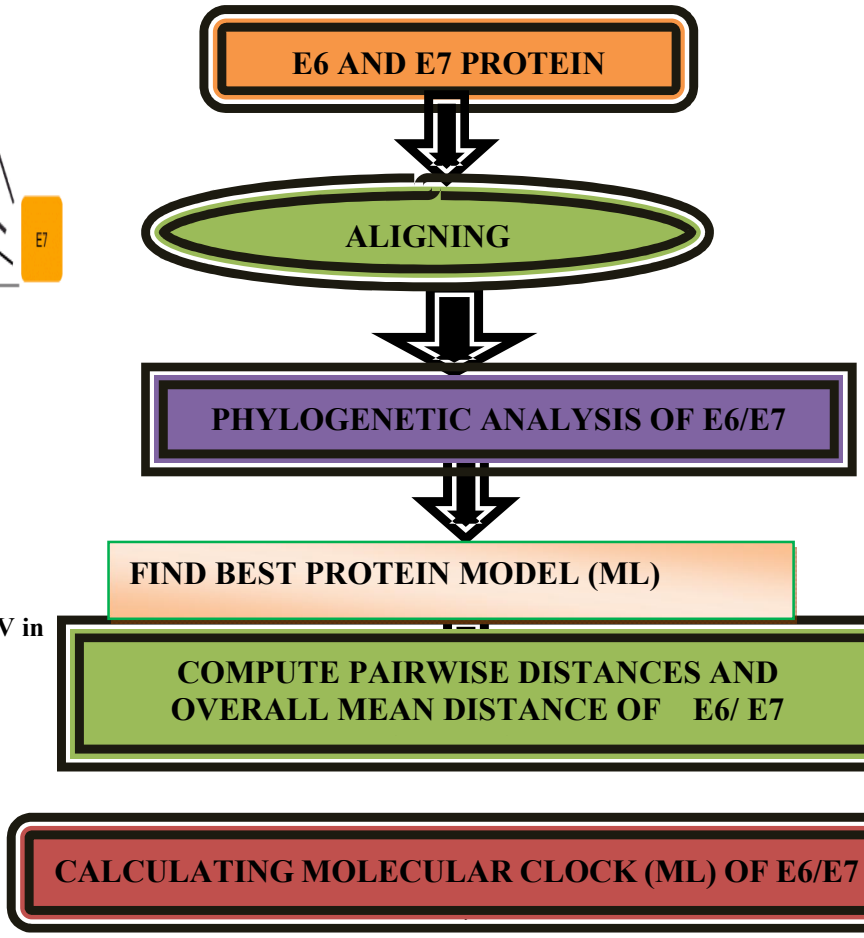
Fig: a (Mechanism of E6/E7 protein of HPV in Carcinogenesis of Cervical Cancer)

IV. TOOLS AND TECHNIQUES

A. Mega Software:

It is online and offline software. But here used offline MEGA5.2 Software to analyze and compare the protein sequences of Human Papillomavirus (HPV). Molecular Evolutionary Genetics Analysis version 5.2 (MEGA) is user friendly software for mining online databases, building sequence alignments and phylogenetic trees, and using methods of evolutionary bioinformatics in basic biology, biomedicine, and evolution. The newest addition in MEGA5.2 is a collection of maximum likelihood (ML) analyses for inferring evolutionary trees, selecting best-fit substitution models (nucleotide or amino acid), inferring ancestral states and sequences (along with probabilities), and estimating evolutionary rates site-by-site. In computer simulation analyses, ML tree inference algorithms in MEGA5.2 compared favorably with other software packages in terms of computational efficiency and the accuracy of the estimates of phylogenetic trees, substitution parameters, and rate variation among sites. This version of MEGA is intended for the Windows platform, and it has been configured for effective use on Mac OS X and Linux desktops. It is available free of charge from http://www.megasoftware.net.

V. METHODOLOGY:



VI. RESULTS AND DISCUSSION:

Sequence alignment and phylogenetic analysis of the E6 and E7 protein of Human papillomavirus of the particular organism of Homo sapiens are:

A. Maximum parsimony analysis

An optimal sequence alignment, critical in the generation of valid phylogenetic trees, was assembled using the amino acid sequence of the E6 protein of HPVs . Owing to the redundancy of the protein sequences, amino acid sequences are more conserved than nucleotide sequences, allowing a more accurate alignment, Although the primary E6 protein sequences are remarkably divergent, all sequenced HPVs contain four

copies of a Cys-X-X-Cys motif spaced at regular and invariant intervals. These motifs and other conserved residues allow an unambiguous alignment of the sequences over 128 amino acids, starting with a conserved aliphatic residue (I, L or V) and ending with a conserved cysteine. Converting the protein sequence into the corresponding nucleic acid sequence resulted in a 384 bp alignment without gaps, which was used for construction of phylogenetic trees using. The branching pattern of the most parsimonious phylogenetic tree served to cluster different HPVs into groups corresponding to their known tissue tropism and oncogenic potential (Fig. b & c) the first major branch (A) consists of viruses that infect the skin. From this branch, one group (I) corresponds to HPV-1, found in skin warts, and a second group (II) corresponds to the EV types. HPV-14, -20, -21 and -25 form one subgroup (IIa), and HPV-5, -47 and -8 form a second subgroup (lib). HPVs in the first subgroup are found only rarely in malignant EV lesions, whereas viruses in the second subgroup have been identified frequently in EV-related skin cancers. The second major branch (B) groups HPVs which primarily infect mucous membranes. An intermediate group (III) consists of HPV-2 and HPV-57. These viruses have been found in papillomas in both the oral mucosa and the skin, causing verruca vulgaris in the latter. The papillomaviruses in the other group (IV) predominantly infect the genital and/or oral mucosae. One subgroup (IVa) comprises HPV-6, -11, -43, -44 and -42, and is associated with benign genital condylomata and low grade cervical intraepithelial neoplasia (CIN). HPV-13, a virus involved in oral focal epithelial hyperplasia (Heck's disease). The viruses in the second subgroup (IVb) are associated with low grade CIN, high grade CIN and invasive anogenital cancer. One branch contains HPV- 16, -31, -33, -35, -51, -52, -56 and -58. A second branch includes HPV-18, -39, -45 and ME180, which have recently been suggested to be associated with more progressive cervical neoplasias.

Fig. b: Alignment of the complete E6 amino acid sequence of 43 HPVs (Homo sapiens)

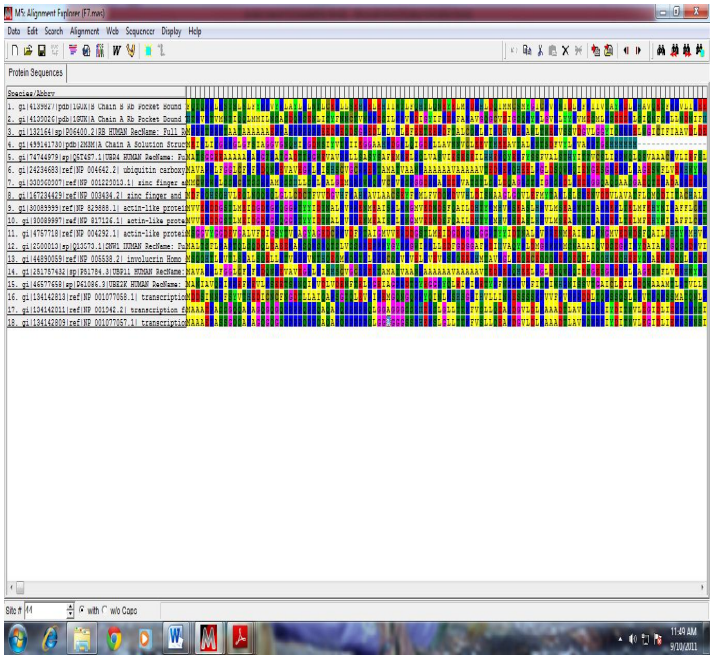


Fig. c: Alignment of the complete E7 amino acid sequence of 18 HPVs (Homo sapiens)

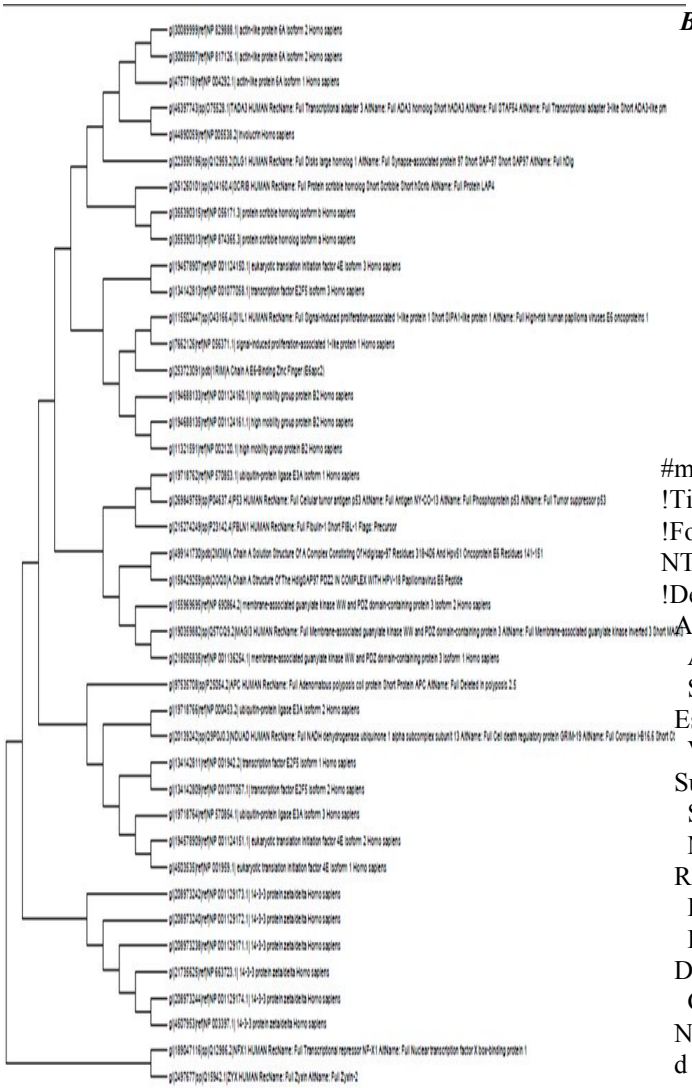


Fig. d: Phylogenetic tree constructed from the alignment of amino acid in the E6 protein of HPVs in 43 protein sequences (Homo sapiens) using maximum parsimony analysis. The corresponding clinical category associated with each HPV type.

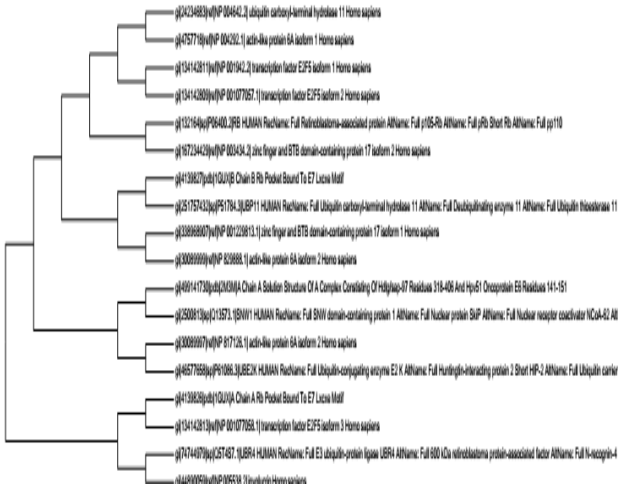


Fig. e: Phylogenetic tree constructed from the alignment of amino acid in the E7 protein of HPVs in 18 protein sequences (Homo sapiens) using maximum parsimony analysis.

B. Distance matrix analysis

To test the validity of the maximum parsimony phylogeny of the HPV E7 sequences, a distance matrix approach (UPGMA) was applied on the same data set, generating the tree shown in Fig. c. In general, the UPGMA tree has the same topology as the PAUP tree (compare Fig. b and Fig.c), with occasional differences in terminal branching. The UPGMA tree clustered HPVs [(20, 21), 14], [(6, 11), (13, 44)] and [16, (35, 31)], whereas the PAUP tree clustered HPVs [(20, 14), 21], [(6, 11), 13, 44] and [(16, 35), 31], respectively. None of these differences had any effect on the classification described for the maximum parsimony tree.

```
#mega
!Title: Phylogenetic Analysis;
!Format DataType=Distance DataFormat=LowerLeft
NTaxa=18;
!Description
 Analysis
  Analysis ---------------------- Distance Estimation
  Scope ------------------------- Pairs of taxa
 Estimate Variance
  Variance Estimation Method ---- None
 Substitution Model
  Substitutions Type ------------ Amino acid
  Model/Method ------------------ Poisson model
 Rates and Patterns
  Rates among Sites ------------- Uniform rates
  Pattern among Lineages -------- Same (Homogeneous)
 Data Subset to Use
  Gaps/Missing Data Treatment --- Complete deletion
 No. of Sites : 97
 d : Estimate
;
[                                                  1]
#gi|4139827|pdb|1GUX|B_Chain_B_Rb_Pocket_Bound_T
o_E7_Lxcxe_Motif
[                                                  2]
#gi|4139826|pdb|1GUX|A_Chain_A_Rb_Pocket_Bound_T
o_E7_Lxcxe_Motif
[3]#gi|132164|sp|P06400.2|RB_HUMAN_RecName:_Full
_Retinoblastoma-associated_protein_AltName:_Full_p105
-Rb_AltName:_Full_pRb_Short_Rb_AltName:_Full_pp11
0
[4]#gi|499141730|pdb|2M3M|A_Chain_A_Solution_Struct
ure_Of_A_Complex_Constisting_Of_Hdlg/sap-97_Residu
es_318-406_And_Hpv51_Oncoprotein_E6_Residues_141-
151
[5]#gi|74744979|sp|Q5T4S7.1|UBR4_HUMAN_RecName:
_Full_E3_ubiquitin
protein_ligase_UBR4_AltName:_Full_600_kDa_retinobla
stoma_protein-associated_factor_AltName:_Full_N-recogn
in-4_AltName:_Full_Retinoblastom1
[                                                  6]
#gi|24234683|ref|NP_004642.2|_ubiquitin_carboxyl-termin
al_hydrolase_11_Homo_sapiens
[7]#gi|338968907|ref|NP_001229813.1|_zinc_finger_and_
BTB_domain-containing_protein_17_isoform_1_Homo_sa
piens
```

[8]#gi|167234429|ref|NP_003434.2|_zinc_finger_and_BTB_domain-containing_protein_17_isoform_2_Homo_sapiens
[9]
#gi|30089999|ref|NP_829888.1|_actin-like_protein_6A_isoform_2_Homo_sapiens
[10]
#gi|30089997|ref|NP_817126.1|_actin-like_protein_6A_isoform_2_Homo_sapiens
[11]
#gi|4757718|ref|NP_004292.1|_actin-like_protein_6A_isoform_1_Homo_sapiens
[12]#gi|2500813|sp|Q13573.1|SNW1_HUMAN_RecName:_Full_SNW_domain-containing_protein_1_AltName:_Full_Nuclear_protein_SkiP_AltName:_Full_Nuclear_receptor_coactivator_NCoA-62_AltName:_Full_Ski-interacting_pron
[13]
#gi|44890059|ref|NP_005538.2|_involucrin_Homo_sapiens
[14]#gi|251757432|sp|P51784.3|UBP11_HUMAN_RecName:_Full_Ubiquitin_carboxyl-terminal_hydrolase_11_AltName:_Full_Deubiquitinating_enzyme_11_AltName:_Full_Ubiquitin_thioesterase_11_AltName:_Full_Ubiquitin-sp1
[15]#gi|46577658|sp|P61086.3|UBE2K_HUMAN_RecName:_Full_Ubiquitin-conjugating_enzyme_E2_K_AltName:_Full_Huntingtin-interacting_protein_2_Short_HIP-2_AltName:_Full_Ubiquitin_carrier_protein_AltName:_Full_Ube
[16]
#gi|134142813|ref|NP_001077058.1|_transcription_factor_E2F5_isoform_3_Homo_sapiens
[17]
#gi|134142811|ref|NP_001942.2|_transcription_factor_E2F5_isoform_1_Homo_sapiens
[18]
#gi|134142809|ref|NP_001077057.1|_transcription_factor_E2F5_isoform_2_Homo_sapiens

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
[1]																		
[2]	2.629																	
[3]	2.495	3.882																
[4]	3.188	3.188	2.495															
[5]	3.188	4.575	2.272	3.476														
[6]	2.965	2.629	2.965	3.188	2.495													
[7]	3.188	2.629	2.965	2.272	3.882	2.272												
[8]	2.965	3.188	3.188	2.783	3.188	3.476	3.476											
[9]	3.188	3.188	2.965	2.629	2.272	2.629	2.783	3.476										
[10]	3.188	3.188	2.965	2.629	2.272	2.629	2.783	3.476	0.000									
[11]	2.965	2.783	2.377	3.188	3.188	2.629	3.476	3.188	2.629	2.629								
[12]	2.495	2.629	2.495	2.965	3.188	1.936	3.882	2.783	2.965	2.965	2.783							
[13]	2.495	2.495	2.090	3.188	3.188	2.495	2.965	2.783	2.629	2.629	2.965	2.090						
[14]	2.965	2.629	2.965	3.188	2.495	0.000	2.272	3.476	2.629	2.629	2.629	1.936	2.495					
[15]	2.965	2.965	2.783	2.965	3.476	2.629	2.965	2.629	2.783	2.783	3.188	2.965	2.965	2.629				
[16]	2.629	2.495	3.882	2.629	2.377	2.965	2.495	2.965	2.377	2.377	2.965	2.629	2.965	2.965	2.629			
[17]	2.629	2.965	2.090	2.377	2.272	3.188	2.495	2.629	2.272	2.272	2.629	2.377	2.629	3.188	2.272	3.188		
[18]	2.629	2.965	2.090	2.377	2.272	3.188	2.495	2.629	2.272	2.272	2.629	2.377	2.629	3.188	2.272	3.188	0.000	

Fig. f: HPV-E7 Protein Compute Data Pairwise Distance.

C. Phylogenetic analyses & of a combined multiple ORF alignment

To investigate whether the phylogenetic trees obtained with the E6 alignment were representative of the molecular evolution of the whole HPV genome, amino acid sequence alignments were constructed for additional ORFs of 18 HPVs for which complete protein sequences were available. A compilation was made of 14 stretches of 20 amino acids or more (totaling 1083 residues) which could be unambiguously aligned (Fig. 6.1 & 6.2). (The data matrix was too large to use UPGMA in the CLUSTAL option of PC/PROTEIN.) The most parsimonious phylogenetic tree for the 3249 bp of the 18 HPVs is shown in Fig. b. The tree had the same overall topology as both the PAUP tree and the UPGMA tree constructed for the E6 protein of the 43sequencses of HPV (Homo sapiens) compare with E7 protein of 18 sequences of HPV (Homo sapiens). The only difference was the position of HPV-51, which was more closely related to HPV-18 and -39 in the 3249 bp tree than in the E6 tree.

D. Comparative Analysis of Different Strains of E6/E7 Proteins of HPV (Homo sapiens) on the Basis of Sequence length and Distance: -

The comparative results of E6 protein of different 43 protein sequences of HPV as well as 18 protein sequences of E7protein different strains of HPV, based on sequence length, and distance. Phylogenetic analysis of E6 protein of 39 strains of HPV revealed genetic divergence of virus proteins into 2 lineages (1 & 2) which further differentiate into 7 small lineage (I to VII) whereas E7 protein of 42 strains of HPV also revealed into 2 lineages (1 & 2) which further differentiated into 8 different lineages (I to VIII). It was found that lineage I, II, III, IV of Phylogenetic tree of E6 proteins appeared as diverged from lineage V (HPV types 92, 96) which itself diverge from lineage VI and VII initially in the evolutionary scale**.** In case of E7 proteins of HPV strains, lineage I, II, III, IV, VIII appeared to diverged from lineage V which initially diverged from VI and VII (Fig c). Conserved amino acids are in bold; the four Cys-X-X-Cys motifs are boxed. The region used to identify nucleic acid sequences for alignment and construction of phylogenetic trees starts at the first conserved aliphatic residue (I, L or V) and ends with the last conserved cysteine (arrows) (fig b).

Further MSA results also revealed that 69.5 %conserved amino acids lineage I (HPV 1, HPV 63) ofE6 protein were fully conserved whereas 18.4 % were highly conserved in.

While in case of lineage I (HPV 1,63) of E7 protein, 52.6% were fully conserved amino acids and 20% were strongly conserved. Lineage II of both E6 (HPV 2, 71, 90, 10, 61, 6, 54, 7, 32) and E7 (2,10, 18, 41, 71, 54, 61, 90) had low percentages of fully conserved and strongly conserved residues. Lineage III (HPV 16, 34, 18, 26, 53) of E6 proteins had30.6 % conserved amino acids and only 17.1 strongly conserved amino acids in the scale of evolution but in case of E7 protein lineage III (HPV 16, 34, 26, 53) it was only 25.2% and 22.4% respectively. Lineage IV of E6(HPV 4, 50, 88, 112, 48, 60, 109, 41) showed low percentage of conserved residue compared to lineage IV of E7 proteins (HPV 4, 88, 60, 112) but a comparable percentage of conserved amino acids were there in case of lineage VI of both E6 and E7 proteins. E7 protein of HPV types 9 and 113 (lineage VII) had 83.8% fully conserved amino acids and 8.6% strongly conserved amino acids whereas E6 protein of lineage VII (HPV9,113,100) had only 49% and 20.9% respectively. Lineage VIII of E7 protein (HPV 101, 103, 108) which was absent in E6 protein as there was no E6 protein in those strains identified, showed 51% fully conserved residues and 21%strongly conserved residues in evolutionary scale (fig .c).

VII. DISCUSSION:

From the comparison of (fig. e), we found that the number of amino acids in E6 protein of different strains of HPV. Though for most HPV, transmission routes, pathogenesis and duration of infection are only poorly understood; phylogenetic analysis of both E6 and E7 proteins of different strains of HPV (Fig. c) revealed that in most of the strains of HPV both E6 and E7 proteins were evolving in a similar manner . As lineage I (HPV 1, HPV 63) of both E6 and E7 proteins showed similar type of evolution in the time scale, this may indicate similar type of pathogenicity of these types mainly associated with common warts as mentioned in genomic database **[16]** and reported by Michael et al **[17]**.The MSA results of different lineages obtained fromE6 and E7 phylogenetic tree reflected the variation indifferent lineages, i.e. in the process of evolution whether there were any insertions or deletions and how many numbers of amino acids were fully conserved or strongly conserved in the protein sequences (fig b). Raioletal also explored the nucleotide variability and phylogeny of the high-risk HPV-31, -33, -35, -52, and -58, in samples from Central Brazil .The different HPV types in different lineages may associate with different types of nongenital and anogenital diseases as mentioned in table 1which give some sort of clinical relatedness to our study. The study provides the genetic diversity of HPV types which may help to understand the oncogenic potential of the virus and to improve management of patients . More than 140 different strains of HPV have been identified however; only genomes of only 42 HPV types have been completely sequenced so far. Sequencing of rest of other strains of HPV may help further evolutionary analysis of HPV.

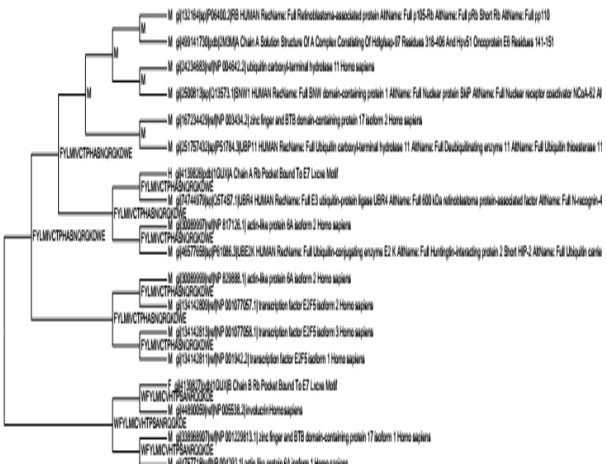


FIG. 8 g hpv e7 protein test molecular clock (ml)

VIII. CONCLUSION

The phylogenetic tree topology obtained on E6 and E7 proteins analysis of 43 sequences of E6 and 18 sequences of E7 protein. HPV strains revealed some sort of divergence among different strains. MSA results among different lineages suggested some variations in amino acids. Besides, also suggested that some conserved residues among divergent lineages of the both the proteins may not be a random process but instead involves mechanisms which lead for causing specific carcinomas. Future investigation into specific protein may provide evidence for understanding co-evolutionary patterns of virus proteins. Also phylogenetic analysis and genetic characterization of other HPV proteins along with E6and E7 may discover some more functional significance of lineage-specific amino acid changes in the internal proteins of HPV.

ACKNOWLEDGMENT

The authors wise to thanks BIFC, D.G.P.G. College, Kanpur funded by Department of Biotechnology, India for providing necessary facilities and financial assistance to pursue the work. My sincere thanks are also due to Indian Science Congress Association for providing, Sir Ashutosh Mukherjee Fellow to Dr. Ashok Kumar Saxena.

REFERENCES

[1] Fiona S. L. Brinkman "Department of Microbiology and Immunology" University of British Columbia Vancouver, British Columbia, Canada .

[2] Detlef D. Leipe "National Center for Biotechnology Information National Library of Medicine National Institutes of Health", Bethesda, Maryland.

[3] Genital HPV Infection — CDC Fact Sheet".Centers for Disease Control and Prevention (CDC). April 10, 2008. Retrieved 13 November 2009.
[4] Kuo, HK; Fujise, K (2011-11-01). "Human papillomavirus and cardiovascular disease among u.s. Women in the national health and nutrition examination survey, 2003 to 2006.". Journal of the American College of Cardiology 58(19)
[5] Gillison ML. "Human papilloma virus-associated head and neck cancer is a distinct epidemiologic, clinical, and molecular entity". Semin Oncol 2004;31:744–54.
[6] Schiffman M, Castle PE "Human papillomavirus: epidemiology and public health". (August 2003). Arch Pathol Lab Med 127 (8):
[7] Walboomers JM, et al. "Human papillomavirus is a necessary cause of invasive cervical cancer worldwide". (1999) J. Pathol. 189 (1):
[8] Goldstein MA, et al. "Case records of the Massachusetts General Hospital. Case 10-2009. A 23-year-old woman with an abnormal Papanicolaou smear". (March 2009). N. Engl. J. Med. 360 (13):
[9] http://www.medscape.org/viewarticle/
[10] Chen, Z.; Schiffman, M.; et al. "Evolution and Taxonomic Classification of Human Papillomavirus 16 (HPV16)-Related Variant Genomes". (2011). PloS ONE 6: 1–16.
[11] Zuna, R. E.; et al. "HPV16 Variant Lineage, Clinical Stage, And Survivalin Women With Invasive Cervical Cancer". Infectious Agents & Cancer **6**: 19–27.
[12] Ganguly, N.; Parihar, S. P. "Human papillomavirus E6 and E7 oncoproteins as risk factors for tumorigenesis". (2009) Journal of biosciences **34** (1): 113–123.
[13] Eligibility Criteria by Topic - American Red Cross.
[14] Chaturvedi, Anil; et al ."Human Papillomavirus and Head and Neck Cancer". In Andrew F. Olshan. Epidemiology, Pathogenesis, and Prevention of Head and Neck Cancer (1st ed.). (March 4, 2010) New York: Springer. ISBN 978-1-4419-1471-2
[15] Tang, S.; Tao, M.; et al. "The E7 Oncoprotein is Translated from Spliced E6*I Transcripts in High-Risk Human Papillomavirus Type 16- or Type 18-Positive Cervical Cancer Cell Lines via Translation Reinitiation".(2006). Journal of Virology **80** (9):42494263.
[16] Münger; et al. "Human papillomavirus immortalization and transformation functions". (2002). Virus research **89** (2): 213–228.
[17] Conway MJ, Alam S, Ryndock EJ et al. "Tissue-spanning redox gradient-dependent assembly of native human papillomavirus type 16 virions". (October 2009). Journal of Virology **83** (20): 10515–26.
[18] Hartwell , LH, L Hood, et al. "Genetics: From Genes to Genomes, 3rd Ed". (2008) McGraw-Hill: New York.
[19] Warnow, T "Computational Methods in Phylogenetics" (2004). Computational Systems Biology Conference, Stanford, CA.
[20] Linder, CR, T Warnow "An overview of phylogeny reconstruction." (2005). In the Handbook of Computational Molecular Biology, Chapman and Hall/CRC Computer & Information Science.
[21] Durbin, R, S Eddy, et al. "Biological Sequence Analysis". (1998). Cambridge University Press: Cambridge.
[22] Linder, CR, T Warnow "An overview of phylogeny reconstruction." (2005). In the Handbook of Computational Molecular Biology, Chapman and Hall/CRC Computer & Information Science.
[23] Liò, P, N Goldman "Models of Molecular Evolution and Phylogeny." (1998). Genome Research, 8:1233-1244.
[24] Li, WH "Molecular Evolution". (1997) Sinauer Associates: Sunderland, MA.
[25] SWOFFORD, D. L. "PAUP: Phylogenetic Analysis Using Parsimony, Version 3.0. Computer program and documentation". (1991). Illinois Natural History Survey, Champaign, Illinois, U.S.A.
[26] WILBUR, W. J. et al. "Rapid similarity searches of nucleic acid and protein data banks". (1983). Proceedings of the National Academy of Sciences, U.S.A. 80, 726 -730.
[27] SOKAL, R. & SNEA'm, P. H. A. "Principles of Numerical Taxonomy". (1963). San Fransisco: Freeman.
[28] HIGGINS, D. G. & SHARP, et al. "Clustal: a package for performing multiple sequence alignment on a microcomputer". (1991). Gene 73, 237-244.
[29] McCULLOUGH, D. W. & McNICOL, et al. "Laryngeal carcinoma associated with human papillomavirus type 16". (1991). Journal of Otolaryn- gology 20, 9799.