ORIGINAL ARTICLE

Comparative Phylogenetic Analysis of E6 and E7 Proteins of Different 42 Strains of HPV

Sangeeta Daf*, Lingaraja Jena, Satish Kumar

Abstract

Simultaneous phylogenetic analysis of E6 and E7 proteins of 42 different HPV strains was carried out in detail. Both E6 and E7 proteins of different HPV strains consistently 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. Dendrogram topologies of E6 and E7 proteins among different HPV strains were very similar which showed that in most of the strains of 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 among different strains. Completion of evolutionary analysis of the E6 and E7 proteins of 42 HPV strains revealed co-dependent evolution of genes with some variations.

Key Words

Phylogenetic analysis, Human Papillomavirus, MSA

Introduction

Human Papillomavirus (HPV) is one of the most common virus groups in the world today affecting the skin and mucosal areas of the body. Over 140 different strains of HPV have been identified (1) which infect different parts of the body. The most visible forms of the virus produce warts (papillomas) on the hands, arms, legs and other areas of the skin. *Table 1* illustrates different diseases associated with various HPV strains. HPV belong to the Papovaviridae family. They consist of 72-capsomere capside containing the viral genome. Capsomers are composed of two structural proteins: the 57 kD late protein L1, which accounts for 80% of the viral particle, and the 43-53 kD minor capside protein L2. The HPV genome, a double-stranded DNA molecule consists of eight kilobase pairs (kbp) nucleotides. Arrangement of the 8-10 open reading frames (ORFs) within the genome is similar in all papillomavirus types and partly overlapping ORFs are arranged on a sole DNA strand. The genome can be divided into three regions: the long control region (LCR) without coding potential; the region of early proteins (E1-E8) & the region of late proteins (L1 & L2) (2). Among all different viral proteins, it is found that E6 and E7 are necessary for HPV-induced malignancy (3). One key activity of E7 is to overcome the pRB tumour suppressor block (4). Binding of E7 to pRB and its related members result in the liberation of E2F transcription factors, which play key role in promoting host cell and viral DNA synthesis. E7 also binds and activates cyclin complexes, such as p33-cyclin dependent kinase (5, 6) which control progression through the cell cycle. HPV E7 proteins of both low and high risk types have an ability to promote unscheduled DNA replication in spinous cells (7-11). E6 protein can overcome the p53 protective control pathways (12), which are important in preventing the genetic damage that may lead to cancer. So our objective was to study the evolution of E6 & E7 proteins among various strains of HPV by comparative phylogenetic analysis whose genomes have been completely sequenced (13). This study thus represents an up-to-date, rigorous and in-depth phylogenetic analysis of E6 & E7 proteins of HPV. The genomic variation of HPV variants obtained from phylogenetic analysis is an essential factor for understanding biological differences of these viruses and contributes further to studies on their infectivity and pathogenicity.

From the Department of * Obstetrics & Gynecology and Bioinformatics Centre, MGIMS, Sevagram (Wardha) Maharashtra-India
Correspondence to : Prof (Dr). Satish Kumar, Biochemistry & Dy Coordinator, Bioinformatics Centre, MGIMS, Sevagram (Wardha).

Table 1. Different Strains of HPV and Diseases Associated with them (16)

HPV Type	Nongenital Cutaneous Disease
1 2 4 26 41	Common warts (verrucae vulgaris)
1 2 4 63	Plantar warts (mymecias)
10 41 49	Flat warts (verrucae plana)
1 2 4 7 10	Butcher's warts (common warts of people who handle meat, poultry, and fish)
2	Mosaic warts
16	Ungual squamous cell carcinoma
2, 50	Epidermodysplasia verruciformis (benign)
9 10 24	Epidermodysplasia verruciformis (malignant or benign)
HPV Type	**Nongenital Mucosal Disease**
6	Respiratory papillomatosis
6, 16, 18	Squamous cell carcinoma of the lung
6	Laryngeal papilloma
16, 18	Laryngeal carcinoma
16, 18	Squamous cell carcinoma of the sinuses
6	Conjunctival papillomas
16	Conjunctival carcinoma
32	Oral focal epithelial hyperplasia (Heck disease)
16, 18	Oral carcinoma
16, 18	Oral leukoplakia
16, 18	Squamous cell carcinoma of the esophagus
HPV Type	**Anogenital Disease**
6, 54	Condylomata acuminate
16, 18, 34	Bowenoid papulosis
16, 18, 34	Bowen disease
6	Giant condylomata (Buschke-Löwenstein tumors)
34 53 61	Unspecified intraepithelial neoplasia
6	Low-grade intraepithelial neoplasia
16 18	High-grade intraepithelial neoplasia
6, 16, 18	Carcinoma of vulva
16	Carcinoma of vagina
16, 18	Cervical Cancer
16 32	Carcinoma of anus
16	Carcinoma in situ of penis (erythroplasia of Queyrat)
16, 18	Carcinoma of penis

Fig 1. Comparative Phylogenetic Trees of E6 & E7 Proteins of Different HPV Strains Constructed by MEGA 4 Software

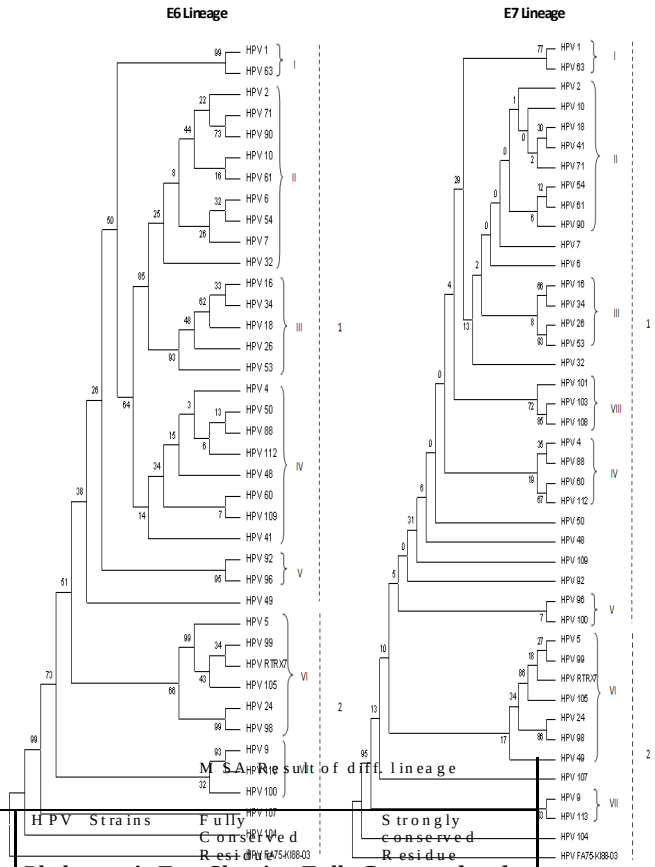

Table 3. MSA Result Showing Different Lineage Group Obtained From Phylogenetic Tree Showing Fully Conserved and Strongly Conserved Residues

E6 Protein Lineage	HPV Strains	Fully Conserved Residue (No. %)	Strongly conserved Residue (No. %)	E7 Protein Lineage	HPV Strains	Fully Conserved Residue (No. %)	Strongly conserved Residue (No. %)
I	1, 63	98 (69.5)	26 (18.4)	I	1, 63	50 (52.6)	19 (20.0)
II	2, 71, 90, 10, 61, 6, 54, 7, 32	24 (14.9)	22 (13.6)	II	2, 10, 18, 41, 71, 54, 61, 90	9 (7.5)	15 (12.6)
III	16, 34, 18, 26, 53	50 (30.6)	28 (17.1)	III	16, 34, 26, 53	27 (25.2)	24 (22.4)
IV	4, 50, 88, 112, 48, 60, 109, 41	18 (11.4)	15 (9.5)	IV	4, 88, 60, 112	30 (29.4)	22 (21.5)
V	92, 96	88 (38.9)	17 (7.5)	V	96, 100	60 (57.6)	15 (14.4)
VI	5, 99, RTRX7, 105, 24, 98	61 (38.8)	26 (16.5)	VI	5, 99, RTRX7, 105, 24, 98, 49	41 (37.6)	16 (14.6)
VII	9, 113, 100	75 (49.0)	32 (20.9)	VII	9, 113	78 (83.8)	8 (8.6)
				VIII	101, 103, 108	51 (51.0)	21 (21.0)

Table 2. Comparative Analysis of Different Strains of E6 & E7 Proteins of HPV on the Basis of Length, MW(kD) & IP (pI)

Strain Name	E6				E7			
	Accession No	Length	Mol. Wt	pI	Accession No	Length	Mol. Wt.	pI
HPV 1	NP_040305	140	16.31	6.80	NP_040307	93	10.50	4.21
HPV 2	NP_077116	159	18.30	8.46	NP_077117	92	10.36	4.56
HPV 4	NP_040889	140	16.48	7.87	NP_040890	100	11.12	4.40
HPV 5	P26556	157	18.08	5.56	CAA52696	103	11.73	4.39
HPV 6	CAS89272	150	17.27	8.20	AAF00065	98	10.90	4.43
HPV 7	NP_041854	150	17.29	8.20	NP_041855	111	12.45	4.80
HPV 9	NP_041861	141	16.39	6.80	NP_041862	93	10.39	4.68
HPV 10	NP_041741	148	17.56	9.05	NP_041742	86	9.54	4.68
HPV 16	NP_041325	158	19.18	9.16	NP_041326	98	11.02	4.20
HPV 18	NP_040310	158	18.87	8.95	NP_040311	105	11.99	4.70
HPV 24	NP_043367	140	16.31	6.79	NP_043368	96	10.71	4.43
HPV 26	NP_041782	159	17.92	8.95	NP_041783	104	11.99	4.08
HPV 32	NP_041801	142	16.63	8.65	NP_041802	104	11.59	4.07
HPV 34	NP_041807	148	17.73	8.79	NP_041808	97	10.98	4.49
HPV 41	NP_040285	156	17.30	7.45	P27556	114	12.80	4.84
HPV 48	NP_043416	142	16.75	8.28	NP_043417	93	10.41	4.93
HPV 49	NP_041832	138	16.20	8.02	NP_041833	103	11.45	4.26
HPV 50	NP_043423	141	16.41	7.82	NP_043424	93	10.51	5.10
HPV 53	AAY69365	154	18.26	9.13	CAK55453	105	12.17	4.47
HPV 54	NP_043288	143	17.13	8.74	AAO15449	95	10.43	4.58
HPV 60	NP_043437	142	16.80	8.37	NP_043438	96	10.68	4.58
HPV 61	NP_043444	146	16.99	7.00	NP_043445	95	10.46	4.40
HPV 63	NP_040896	141	16.31	8.08	NP_040897	88	9.47	4.23
HPV 71	AAQ95184	156	17.90	7.96	NP_597935	89	10.06	4.34
HPV 88	YP_001672008	142	16.73	8.14	YP_001672009	98	10.89	4.54
HPV 90	AAL14204	148	17.17	6.79	AAL14205	98	10.94	4.58
HPV 92	NP_775305	138	15.80	6.29	NP_775306	91	10.11	4.35
HPV 96	NP_932319	225	26.04	9.07	NP_932320	99	10.03	4.38
HPV 98	YP_002922749	153	17.72	6.79	YP_002922750	99	11.03	4.38
HPV 99	YP_002922756	155	17.64	5.71	YP_002922757	103	11.58	4.39
HPV 100	YP_002922762	152	18.18	7.93	YP_002922763	100	11.19	4.38
HPV 101	**Not Available**				YP_656499	98	10.74	4.88
HPV 103	**Not Available**				YP_656493	100	11.54	4.94
HPV 104	YP_002922922	138	16.43	6.88	YP_002922923	104	11.67	4.35
HPV 105	YP_002922769	155	17.81	5.35	YP_002922770	101	11.26	4.22
HPV 107	ABN79867	140	16.55	5.51	ABN79868	102	11.58	4.51
HPV 108	**Not Available**				YP_002647034	99	11.16	4.89
HPV 109	YP_002756538	140	16.05	7.35	YP_002756539	96	10.69	4.75
HPV 112	YP_002756545	139	16.39	8.29	YP_002756546	97	10.83	4.47
HPV 113	YP_002922775	149	17.42	6.71	YP_002922776	92	10.50	4.60
HPV RTRX7	NP_847990	157	18.16	6.29	NP_847991	103	11.49	4.21
HPV FA75-KI88-03	YP_002235531	138	16.41	6.88	YP_002235532	104	11.68	4.37

Materials & Methods

Data collection: There are 42 strains of HPV whose genomes are completely sequenced and available at National Centre for Biotechnology Information (NCBI). So all E6 & E7 protein sequences of HPV strains were downloaded from NCBI in fasta format and then subjected to ExPASy ProtParam (14) tool for calculating their molecular weight and theoretical Isoelectric Point (pI). Then all the results were put in tabular format separately for E6 and E7 protein sequences.

Phylogenetic analysis: A multiple sequence alignment (MSA) of E6 & E7 protein sequence of 42 different HPV strains was performed using ClustalW module of MEGA 4 (Molecular Evolutionary Genetics Analysis) software (15). The gap penalty of 10 was set for both pair wise as well as MSA where as gap extension penalty of 0.1 and 0.2 was set for pair wise and MSA respectively. BLOSUM was selected as protein weighted matrix for performing MSA. Then mega (.meg) file were exported from MSA results. The corresponding mega files were then subjected to phylogenetic tree construction using maximum parsimony method showing bootstrap value at nodes as all the strains are closely related to each other and there are strong sequence similarities among themselves.

MSA for different lineages: Different sequence groups (the lineages obtained from phylogenetic tree construction result) were subjected to ClustalX for obtaining MSA and identifying the variations among closely related lineages.

Results

Table 2 shows the comparative results of E6 protein of 39 different strains of HPV as well as E7 protein of

42 different strains of HPV, based on sequence length, Molecular weight and Isoelectric Point (pI).

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) (*Fig 1*). 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 1*).

Further MSA results also revealed that 69.5 % conserved amino acids lineage I (HPV 1, HPV 63) of E6 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 (*Table 3*). Lineage III (HPV 16, 34, 18, 26, 53) of E6 proteins had 30.6 % 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 (HPV 9,113,100) 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 (*Table 3*).

Discussions

From the comparison of *table 2*, we found that the number of amino acids in E6 protein of different strains of HPV varies from 138 to 159 having molecular weight in the range of 15.80 kD to 19.18 kD and pI from 5.51 to 9.16 except HPV type 96 which has longest E6 protein (225 amino acids, Mol Wt. 26.04 kD & pI 9.07). However, the number of amino acids in E7 protein of different strains of HPV varies from 86 to 104 having molecular weight in the range of 9.47 KDa to 12.8 KDa and pI 4.07 to 4.80.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 1*) revealed that in most of the strains of HPV both E6 and E7 proteins were evolving in a similar manner (*Table 3*). 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 from E6 and E7 phylogenetic tree reflected the variation in different 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 (*Table 3*). Raiol *et al* also explored the nucleotide variability and phylogeny of the high-risk HPV-31, -33, -35, -52, and -58, in samples from Central Brazil (18).

The different HPV types in different lineages may associate with different types of nongenital and anogenital diseases as mentioned in *table 1* which 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.

Conclusion

The phylogenetic tree topology obtained on E6 and E7 proteins analysis of 42 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 E6 and E7 may discover some more functional significance of lineage-specific amino acid changes in the internal proteins of HPV. Also further studies on E6 and E7 proteins especially their three dimensional structures may lead to design efficient genome based drugs and vaccines for different carcinomas of HPV.

References

1. Peter M, Howley B. Warts, Cancer And Ubiquitylation: Lessons From The Papillomaviruses. Transactions of The American Clinical and Climatological Association, Massachusetts 2006; Vol. 117.pp.35-57.
2. Baker CC, Phelps WC, Lindgren V, Braun MJ, Gonda MA, Howley M. Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines. *J Virol* 1987; 61:962-67.
3. Jason J. Chen, Cell cycle checkpoints, genomic instability, and apoptosis in cells expressing papillomavirus oncogenes. (Cited 2009 June, 28). Available from: http://www.umassmed.edu/faculty/show.cfm?start=0&faculty=356
4. Dyson N, Howley PM, Munger K, *et al*. The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. *Science* 1989; 243:934-36.
5. Franco EL. The sexually transmitted disease model for cervical cancer: incoherent epidemiologic findings and the role of misclassification of human papillomavirus infection. *Epidemiology* 1991; 2:98-106.
6. Tommasino M, Adamczewski JP, Carlotti F, *et al*. HPV16 E7 protein associates with the protein kinase p33CDK2 and cyclin A. *Oncogene* 1993; 8:195-202.
7. Cheng S, Schmidt-Grimminger DC, Murant T, *et al*. Differentiation-dependent up-regulation of the human papillomavirus E7 gene reactivates cellular DNA replication in suprabasal-differentiated keratinocytes. *Genes & Dev* 1995; 9:2335-49.
8. Jian Y, Schmidt-Grimminger D-C, Chien W-M, *et al*. Post-transcriptional induction of p21cip1 protein by human papillomavirus E7 inhibits unscheduled DNA synthesis reactivated in differentiated keratinocytes. *Oncogene* 1998; 17:2027-38.
9. Schmidt-Grimminger D-C, Wu X, Jian Y, *et al*. Post-transcriptional induction of p21cip1 protein in condylomata and dysplasias is inversely related to human papillomavirus activities. *Am J Pathol* 1998; 152:1015-24.
10. Steenbergen RDM, Parker JN, Isern S, *et al*. Viral E6-E7 transcription in the basal layer of organotypic cultures without apparent p21cip1 protein precedes immortalization of HPV 16 and HPV 18 transfected human keratinocytes. *J Virol* 1998; 72:749-57.
11. Demeter LM, Stoler MH, Broker TR, *et al*. Induction of proliferating cell nuclear antigen in differentiated keratinocytes of human papillomavirus-infected lesions. *Hum Pathol* 1994; 25:343-48.
12. Werness BA, Levine AJ, Howley PM. Association of human papillomavirus types 16 and 18 E6 proteins with p53. *Science* 1990; 248:76-79.
13. Daf S, Jena L, Kumar S. Genome sequencing of HPV - A hope to overcome cervical cancer. *J MGIMS* 2008; 13: 1-8.
14. ExPASy - ProtParam tool (Cited 2009 July 27). Available from: http://www.expasy.ch/tools/protparam.html
15. Kumar S, Dudley J, Nei M & Tamura K . MEGA (Molecular Evolutionary Genetics Analysis) (Cited 2009 July 21). Available from: http://www.megasoftware.net/
16. Viruses Genomes - Human Papillomavirus. (Cited 2009 July 15). Available from: http://www.ebi.ac.uk/2can/genomes/viruses/Human_papillomavirus.html
17. Michael KM, Waterboer T, Sehr P, *et al*. Seroprevalence of 34 Human Papillomavirus Types in the German General Population. *PLoS Pathog* 2008; 4(6): e1000091.
18. Raiol T, Wyant PS, de Amorim RM, *et al*. Genetic variability and phylogeny of the high-risk HPV-31, -33, -35, -52, and -58 in central Brazil. *J Med Virol* 2009; 81(4):685-92.

Journal is indexed in Excerpta Medica / EMBASE, MedLine Locator Plus, Ulrich Periodical Dictionary, IndexCopernicus International, Directory of Open Access Journals (DOAJ), OpenMed@NIC, Indian Science Abstract, SCOPUS, SCOLOAR, Genamics JournalSeek, Biology & Medicine Online Journals, UONLe-Journal List, SC Imago Journal Ranking, EBSCO Publishing's Electronic Database, CIS Indexing & Abstracts, Electronic Journal Index (SJSU), Global Health, CAB Abstracts Databases, Open J Gate Informatics, ProQuest Health & Medical Database & Many International E - Indexing Agencies & Libraries