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

<table>
<thead>
<tr>
<th>HPV Type</th>
<th>Nongenital Cutaneous Disease</th>
<th>Nongenital Mucosal Disease</th>
<th>Anogenital Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 4 6 3</td>
<td>Plantar warts (verrucae vulgaris)</td>
<td>Respiratory papillomatosis</td>
<td>Condylomata acuminata</td>
</tr>
<tr>
<td>10 41 49</td>
<td>Flat warts (verrucae plana)</td>
<td>Squamous cell carcinoma of the lung</td>
<td>Low-grade intraepithelial neoplasia</td>
</tr>
<tr>
<td>1 2 4 7 10</td>
<td>Butcher’s warts (common warts of people who handle meat, poultry, and fish)</td>
<td>Laryngeal papilloma</td>
<td>High-grade intraepithelial neoplasia</td>
</tr>
<tr>
<td>2</td>
<td>Mosaic warts</td>
<td>Bowenoid papulosis</td>
<td>Unspecified intraepithelial neoplasia</td>
</tr>
<tr>
<td>16</td>
<td>Ungual squamous cell carcinoma</td>
<td>Carcinoma of vagina</td>
<td>Carcinoma of vulva</td>
</tr>
<tr>
<td>2, 50</td>
<td>Epidermodysplasia verruciformis (benign)</td>
<td>Carcinoma in situ of vagina</td>
<td>Carcinoma in situ of penis (erythroplasia of Queyrat)</td>
</tr>
<tr>
<td>9 10 24</td>
<td>Epidermodysplasia verruciformis (malignant or benign)</td>
<td>Carcinoma of vagina</td>
<td>Carcinoma of penis</td>
</tr>
</tbody>
</table>

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

- **E6 Protein**
  - **E7 Protein**
  - **Lineage**
  - **HPV Strains**
    - **Fully Conserved Residue**
    - **Strongly Conserved Residue**

Fig 1. Comparative Phylogenetic Trees of E6 & E7 Proteins of Different HPV Strains Constructed by MEGA 4 Software
### 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 39 different strains of HPV on the basis of Length, MW(kD) & IP (pI).
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 & pl 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


3. Jason J. Chen, Cell cycle checkpoints, genomic instability, and apoptosis in cells expressing papillomavirus oncogenes. (Cited 2009 June, 28). Available from: h t t p : / / w w w . u m a s s m e d . e d u / f a c u l t y / show.cfm?start=0&faculty=356


14. ExPASy - ProtParam tool (Cited 2009 July 27). Available from: h t t p : / / w w w . e x p a s y . c h / t o o l s / protparam.html


