Identification and characterization of two novel human papillomaviruses (HPVs) by overlapping PCR: HPV102 and HPV106

Zigui Chen,1 Mark Schiffman,2 Rolando Herrero3 and Robert D. Burk1,4

1Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, NY, USA
2Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, US Department of Health and Human Services, Bethesda, MD, USA
3Proyecto Epidemiológico Guanacaste, Costa Rican Foundation for Health Sciences, San José, Costa Rica
4Department of Pediatrics, Department of Epidemiology and Population Health, Department of Obstetrics, Gynecology and Woman's Health and Albert Einstein Cancer Center, Albert Einstein College of Medicine, Bronx, NY, USA

Complete genomes of HPV102 (8078 bp) and HPV106 (8035 bp) were PCR amplified and cloned from cervicovaginal cells of a 49-year-old Hispanic female with reactive changes on her Pap test and a 42-year-old Hispanic female with a Pap test diagnosis of atypical squamous cells of unknown significance (ASCUS), respectively. The nucleotide sequence similarity of the complete L1 open reading frame (ORF) determined that HPV102 and HPV106 are most closely related to HPV83 (84.1% identity) and HPV90 (83.5% identity), respectively, placing them in the genital HPV groups, papillomaviruses species a3 and a15. HPV102 and HPV106 contain five early genes (E6, E7, E1, E2, and E4) and two late genes (L2 and L1), and both lack an E5 ORF. On the basis of phylogenetic analyses and available clinical information, these two novel HPV types expand the heterogeneity of HPVs detected in the lower genital tract.

Human papillomaviruses (HPVs) are a family of related viruses with circular double-stranded DNA genomes of approximately 8 kb that contain three general regions. HPVs are highly species-specific pathogens that usually cause benign tumours (warts, papillomas) in the skin and mucosal epithelia of human beings. Persistent infection of the genital tract by high-risk HPV types (e.g. HPV16 and HPV18) is the most important risk factor for development of genital malignancies such as squamous cell carcinoma (SCC) or, less commonly, adenocarcinoma of the cervix (Moscicki et al., 2006). Among over 100 characterized HPVs, approximately 60 predominantly infect the genital tract epithelia. For a papillomavirus to be recognized as a distinct type, the full genome should be cloned and the sequence of the L1 gene should be no more than 90% similar to previously typed PVs (de Villiers et al., 2004).

A phylogenetic tree of the alpha PVs based upon full genome alignments suggested that at least three ancestral PVs are responsible for the current heterogeneous group of genital HPV genomes (alpha PVs) (Schiffman et al., 2005). Carcinogenic types, as defined at a recent workshop of the International Agency for Research on Cancer (IARC) (Cogliano et al., 2005), include types limited to the alpha papillomavirus species α9, α11, α7, α5 and α6 that are derived from a common ancestor (Schiffman et al., 2005), whereas genital HPV types derived from the other two ancestors appear to be non-pathogenic or have low oncogenic potential. In this report, two novel HPV types, HPV102 and HPV106, were identified, cloned and characterized. The L1 nucleotide sequence similarity by pairwise alignment indicated that HPV102 and HPV106 are closely related to HPV83 (genital HPV species α3) and HPV90 (genital HPV species α15), respectively.

Clinical specimens containing HPV102 and HPV106 originated from a 49-year-old Hispanic female with reactive changes on her Pap test and a 42-year-old Hispanic female with a Pap test diagnosis of atypical squamous cells of unknown significance (ASCUS), respectively. They were participating in a population-based HPV study in Costa Rica (Herrero et al., 2005). HPV106 DNA was initially identified by PCR amplification with MY09/11 primers (Castle et al., 2002), but did not hybridize to over 40 type-specific probes used in the HPV typing assays (Herrero et al., 2005). HPV102 DNA was not amplified by MY09/MY11 PCR, but was positive by FAP primers within the L1 ORF (Forslund et al., 1999). The 455 bp

Location of the predicted ORFs of HPV102 and HPV106 and sequence of the primers used are available as supplementary material with the online version of this paper.
fragment amplified by MY09/MY11 primers from HPV106 and the 480 bp fragment amplified by the FAP primers from HPV102 were sequenced. A BLAST search against GenBank revealed that sequences of these MY and FAP fragments were 83 % related to HPV90 and 83 % related to HPV83, respectively. Thereafter, type-specific primer sets were designed based on available sequences and a consensus region within the E1 ORF for amplification of the complete genome in two fragments for HPV102 (4007 bp and 4453 bp) and three fragments for HPV106 (3980, 4006 and 412 bp). Oligonucleotide primer sequences used in this study are available from the authors. For overlapping PCR, an equal mixture of the AmpliTaq Gold Taq DNA polymerase (Applied Biosystems) and Pwo Platinum Taq DNA Polymerase (High Fidelity, Invitrogen) were utilized as previously described (Terai & Burk, 2001).

For HPV cloning and sequencing, each PCR product was purified from agarose gel (Qiagen gel extraction kit) after confirmation of appropriate size, ligated into the pGEM-T easy vector (Promega) and sequenced by the Einstein Sequencing Facility, New York. Subsequent sequencing was performed using primer walking. The complete genomic sequences were determined from combined sets of sequences from overlapping amplification products. DNA clones were submitted to the Human Papillomavirus Reference Laboratory (Heidelberg, Germany) for official designation. The genome sequences of HPV102 and HPV106 were submitted to the NCBI/GenBank database (accession nos: HPV102, DQ080083; HPV106, DQ080082).

Phylogenetic trees were constructed to assess the position of HPV102 and HPV106 in relationship to other HPV types, MrBayes v3.0b4 (Huelsenbeck & Ronquist, 2001) with 10⁷ cycles for the Markov chain Monte Carlo (MCMC) algorithm was used to generate trees from the alignments of concatenated amino acid and nucleotide sequences of early (E6, E7, E1, E2 and E4) and late (L2 and L1) ORFs of concatenated amino acid and nucleotide sequences of HPV102, DQ080083; HPV106, DQ080082). The genome sequences of HPV102 and HPV106 were submitted to the NCBI/GenBank database (accession nos: HPV102, DQ080083; HPV106, DQ080082).

Table 1. Nucleotide sequence pairwise similarity (%) of HPV102 and HPV106 ORFs and URRs with related HPV types

<table>
<thead>
<tr>
<th>Novel types</th>
<th>Pairwise similarity of each ORF and the URR (%)</th>
<th>L1:E1 ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Related types (species)</td>
<td>E6</td>
<td>E7</td>
</tr>
<tr>
<td>HPV102</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPV83 (z3)</td>
<td>85.0</td>
<td>87.3</td>
</tr>
<tr>
<td>HPV89 (z3)</td>
<td>72.7</td>
<td>72.0</td>
</tr>
<tr>
<td>HPV61 (z3)</td>
<td>66.7</td>
<td>73.4</td>
</tr>
<tr>
<td>HPV72 (z3)</td>
<td>69.8</td>
<td>71.2</td>
</tr>
<tr>
<td>HPV106</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPV90 (z15)</td>
<td>78.7</td>
<td>85.4</td>
</tr>
<tr>
<td>HPV71 (z15)</td>
<td>68.2</td>
<td>78.7</td>
</tr>
</tbody>
</table>
The genomes of HPV102 and HPV106 contain seven ORFs, potentially encoding five early genes (E6, E7, E1, E2 and E4) and two late genes (L2 and L1) (Supplementary Figure S1). None of the small ORFs within the genomes of these two novel types shows significant homology with known E5 proteins (data not shown).

The first ATG of the E6 ORF was defined as the first position of the complete genomes of HPV102 and HPV106. The putative E6 ORFs of HPV102 and HPV106 contain two zinc-binding domains CxxC(x)29CxxC, separated by 36 aa. The putative E7 ORFs of HPV102 and HPV106 contain one conserved zinc binding domain, CxxC(x)29CxxC, and a motif (LxCxE) for binding to the pRB protein. The ATP-binding site of the ATP-dependent helicase (GPSDTGKS for HPV102, GPADTGKS for HPV106) is conserved in the carboxy-terminal region of E1. The E4 ORF contains a start codon and overlaps the E2 ORF. A polyadenylation consensus sequence (AATAAA) for processing of early viral mRNA transcripts is present at the beginning of the L2 gene. The major (L1) and minor (L2) capsid proteins of both types show a nuclear localization signal at their 3′ end.

The upstream regulatory regions (URRs) located between the stop codon of L1 and the first start codon of the E6 ORF of HPV102 and HPV106 consist of 782 and 733 bp, respectively, and contain many cis-acting regulatory sequences that control viral transcription and replication. Analysis of the HPV102 URR revealed four typical palindromic E2-binding sites [ACC(N)6GGT] at nt 7569–7580, 7608–7619, 7934–7945 and 8014–8025. Analysis also revealed three typical palindromic E2-binding sites (nt 7592–7603, 7939–7950 and 8015–8026) and a highly-related E2-binding site (nt 7872–7883) within the URR of HPV106. Multiple binding sites for transcriptional regulatory factors such as AP-1 (Chen et al., 1990), NF-1 (Apt et al., 1993), SP-1 (Gloss & Bernard, 1990), transcriptional enhancer factors (TEF)-1 (Ishiji et al., 1992) and YY-1 (Dong et al., 1994) are also present within the URR regions.

To investigate the relationships between these two novel types and related HPV genomes, phylogenetic trees were constructed and compared based on the concatenated amino acid and nucleotide sequences of early (E6, E7, E1, E2 and E4) and late (L2 and L1) ORFs. Representative HPV types within species z2, z3, z4 and z15 were used to calculate Bayesian trees (Fig. 1). HPV32 and HPV42 within the species z1 were set as the ‘referent outgroup’. Both early and late trees indicated that HPV102 and HPV106 were most closely related to HPV83 and HPV90, and were placed into genital HPV species z3 and z15, respectively. The MP and NJ trees also demonstrated stable and consistent topological positions for HPV102 and HPV106 like that in the Bayesian tree (data not shown). However, the ILD test detected significant incongruent evolutionary patterns between the ‘early ORF’ and ‘late ORF’ trees, as shown in Fig. 1 (P=0.007). In the ‘early ORF’ tree, the z3 species forms a sister group to the clade consisting of species z2, z4 and z15, while the ‘late ORF’ tree indicated the close relationship between species z3 and z4. The incongruence in the genital HPV phylogeny could reflect different evolutionary histories or evolutionary pressures among different genes (Narechania et al., 2005).

Novel HPV types continue to emerge, and the clinical significance of each type needs to be evaluated empirically. Among 8513 women participating in a population-based HPV study in Costa Rica (Herrero et al., 2005), at base-line 26.4 % of the study participants had detectable HPV. HPV16 (z9) was the most common oncogenic type (3.6 % prevalence), while HPV61 (z3) and HPV71 (z15) were the most common non-oncogenic types (2.4 and 2.4 % prevalence, respectively). However, HPV102 and HPV106 appear to be rare in the female genital tract. HPV102 was identified in 4/7125 samples, and HPV106 was identified in

![Fig. 1. Bayesian trees inferred from the concatenated amino acid and nucleotide sequences of early (E6, E7, E1, E2 and E4) and late (L2 and L1) ORFs. HPV102 and HPV106, highlighted in bold, are most closely related to HPV83 (84.1 % identity) and HPV90 (83.5 % identity), and are placed in the genital HPV genome homology species z3 and z15, respectively.](https://example.com/fig1.png)
Min 1/1993 samples. However, these estimates of prevalence are the lower limit of detectability, since the PCR systems have not been optimized for detection of HPV102 and HPV106.

On the basis of phylogenetic analyses and available clinical information, these two novel HPV types expand the heterogeneity of HPV's detected in the lower genital tract.

Acknowledgements

We acknowledge Anne Breheny, Amy Razukiewicz and Andrew Prior for performing HPV DNA analyses. This work was supported in part by Public Health Service award CA78527 from the National Cancer Institute.

References


http://vir.sgmjournals.org