Characterization of a novel genital human papillomavirus by overlapping PCR: candHPV86 identified in cervicovaginal cells of a woman with cervical neoplasia

Masanori Terai and Robert D. Burk

Departments of Microbiology & Immunology, Pediatrics, and Epidemiology & Social Medicine, Comprehensive Cancer Center, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York, 10461, USA

A novel human papillomavirus (HPV), candHPV86, was cloned and characterized from cervicovaginal cells obtained from a 37-year-old Hispanic woman with cervical intraepithelial neoplasia grade 1 (CIN1) using an overlapping PCR technique. Primers were designed by phylogenetic alignment of closely related HPV genomes using the L1 fragment sequence amplified by GP5+/6+. The 7983 bp complete nucleotide sequence of the HPV genome was determined by sequence walking. A basic local alignment sequence tool (BLAST) homology search using the L1 open reading frame demonstrated that this HPV was most closely related to HPVHAN2294 (GenBank, AJ400628; 86% homology) and HPV84 (84% homology). candHPV86 was placed in the HPV genome homology group A3 by phylogenetic analyses. The overlapping PCR technique is applicable for characterizing the complete spectrum and variation of HPVs in a population.

The papillomaviruses are a heterogeneous group of viruses with 8 kb closed circular, double-stranded DNA genomes. These viruses infect humans as well as numerous and diverse animal species (Sundberg et al., 1994). Papillomaviruses are classified by the homology of their genome. A cloned human papillomavirus (HPV) genome whose L1 open reading frame (ORF) displays less than 90% similarity to previously designated types is defined as a novel type (Delius et al., 1998). Over 100 putative HPV types have been described, including 85 cloned and officially designated, and others identified from the sequence of PCR products. Genital HPVs are sexually transmitted and a subset of these are now considered to be the aetiological agents of cervical neoplasia and cancer throughout the world (Bosch et al., 1995).

We have isolated and characterized a novel genital HPV type that was amplified and cloned from cervicovaginal cells obtained from a 37-year-old Hispanic woman with cervical intraepithelial neoplasia grade 1 (CIN1) by the overlapping PCR method (Terai & Burk, 2001). The overlapping PCR method is a technique which facilitates the complete characterization of HPV DNA from specimens with low copy numbers. Here, we describe candHPV86, a novel HPV genome, and present the complete nucleotide sequence, genome organization, predicted proteins and phylogenetic analyses.

The clinical sample was found to be HPV DNA-positive by Psil cleavage and Southern blot hybridization but not by MY09/MY11 amplification during the course of a clinical study (Burk et al., 1996; Ho et al., 1998). To further characterize this HPV genome, amplification with the GP5+/6+ primers was performed and the PCR product was sequenced (de Roda Husman et al., 1995). The initial overlapping PCR primers were designed by alignment of closely related HPV genomes determined by basic local alignment sequence tool (BLAST) analysis using the sequence of the partial L1 region. Additional primers were designed to amplify the entire genome in two large fragments. Two sets of primers were designed (Fig. 1a): for amplification of fragment no. 1, forward primer 1-F (5’TTTTACTATTAGTGCCGCTAC 3’) and reverse primer 1-R (5’GCATCTAAACGATCGGCTAG 3’); and for amplification of fragment no. 2, forward primer, 2-F (5’GTATGGCAATACCGAGGTGG 3’) and reverse primer, 2-R (5’AGAGGGTCTATTCAGAGG 3’). Amplification was performed using either Gold Taq DNA polymerase (Perkin-Elmer Applied Biosystems) or an equal mixture of Gold Taq and Poo DNA polymerase (Platinum Taq DNA Polymerase High Fidelity, Gibco-BRL). Poo polymerase has an inherent 3’→5’ exonuclease proofreading activity. The PCR products were separated by electrophoresis in agarose gels, stained with ethidium bromide, and visualized under ultraviolet (UV) illumination. After confirmation of appropriate product size, each PCR product was purified (Qiagen Gel Extraction Kit, Qiagen) and ligated into the pGEM-T Easy vector (Promega) according to the manufacturer’s instructions. To determine the
nucleotide sequence, each DNA insert was initially sequenced using SP6 and T7 primers flanking the HPV insert. Additional primers were designed by sequence walking (Delius & Hofmann, 1994). Sequencing was performed on an ABI Prism Model 377 automated sequencer (Perkin-Elmer Applied Biosystems) in the Einstein DNA sequencing core facility. The sequence of the overlapping fragments was assembled manually and confirmed by sequencing the complementary strand. Several additional primers were designed and used to clarify sequence ambiguities. Once assembled, the sequence was analysed for homology to other HPVs using the BLAST software (Altschul et al., 1997). The same software was used to determine protein sequence homologies. Phylogenetic trees were created using HPV published sequences available from the Human Papillomaviruses Compendiums On Line (http://www.stdgen.lanl.gov/stdgen/virus/hpv/compendium/htdocs) and GenBank. Phylogenetic trees were derived from individual ORFs and long control regions (LCRs) to determine the relationship of candHPV86 to the available HPV sequences using public domain software (Higgins & Sharp, 1988).

Fig. 1(a) displays the organization of the three overlapping PCR products covering the whole candHPV86 genome. The sizes of fragments no. 1, no. 2 and the GP5+6+ fragment were 3-9 kb, 5-8 kb and 140 bp, respectively. The complete nucleotide sequence is shown in Fig. 1(b). The assembled sequence of the virus genome revealed a total size of 7983 bp with a G+C content of 45-92%. The sequence is available in

Table 1. Characterization of the candHPV86 genome

(a) Location of predicted ORFs and size of putative proteins

<table>
<thead>
<tr>
<th>ORF</th>
<th>Start position</th>
<th>First ATG</th>
<th>Stop codon</th>
<th>Length of protein-coding sequence (bp)</th>
<th>Amino acids</th>
<th>Predicted molecular mass of protein (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E6</td>
<td>7957</td>
<td>1</td>
<td>447</td>
<td>444</td>
<td>148</td>
<td>17-2</td>
</tr>
<tr>
<td>E7</td>
<td>417</td>
<td>423</td>
<td>707</td>
<td>282</td>
<td>94</td>
<td>10-4</td>
</tr>
<tr>
<td>E1</td>
<td>649</td>
<td>709</td>
<td>2667</td>
<td>1956</td>
<td>652</td>
<td>73-1</td>
</tr>
<tr>
<td>E2</td>
<td>2573</td>
<td>2609</td>
<td>3733</td>
<td>1122</td>
<td>374</td>
<td>42-4</td>
</tr>
<tr>
<td>E4</td>
<td>3156</td>
<td>3174</td>
<td>3500</td>
<td>324</td>
<td>108</td>
<td>12-2</td>
</tr>
<tr>
<td>L2</td>
<td>4229</td>
<td>4241</td>
<td>5677</td>
<td>1434</td>
<td>478</td>
<td>50-8</td>
</tr>
<tr>
<td>L1</td>
<td>5514</td>
<td>5655</td>
<td>7172</td>
<td>1515</td>
<td>505</td>
<td>50-8</td>
</tr>
</tbody>
</table>

(b) Homology (%) of candHPV86 amino acid sequences with related HPVs

<table>
<thead>
<tr>
<th>candHPV86</th>
<th>Group*</th>
<th>E6</th>
<th>E7</th>
<th>E1</th>
<th>E2</th>
<th>E4</th>
<th>L2</th>
<th>L1</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPVHAN2294</td>
<td>A3</td>
<td>80-3</td>
<td>86-5</td>
<td>88-9</td>
<td>80-0</td>
<td>75-7</td>
<td>85-9</td>
<td>90-9</td>
</tr>
<tr>
<td>HPV84</td>
<td>A3</td>
<td>74-1</td>
<td>77-1</td>
<td>85-9</td>
<td>74-7</td>
<td>67-3</td>
<td>85-8</td>
<td>85-9</td>
</tr>
<tr>
<td>HPV61</td>
<td>A3</td>
<td>66-0</td>
<td>58-9</td>
<td>70-7</td>
<td>59-3</td>
<td>54-7</td>
<td>61-2</td>
<td>78-3</td>
</tr>
<tr>
<td>HPV72</td>
<td>A3</td>
<td>59-5</td>
<td>58-6</td>
<td>70-8</td>
<td>61-3</td>
<td>51-9</td>
<td>62-2</td>
<td>78-0</td>
</tr>
<tr>
<td>HPV83</td>
<td>A3</td>
<td>57-4</td>
<td>61-9</td>
<td>72-7</td>
<td>58-9</td>
<td>46-6</td>
<td>72-2</td>
<td>78-5</td>
</tr>
</tbody>
</table>

* See Chan et al. (1995).
Fig. 1. The genome of candHPV86. (a) The organization of the three cloned overlapping PCR products: plasmid no. 1, plasmid no. 2 and the GP5\(^{\text{&\textregistered}}\)/6\(^{\text{&\textregistered}}\) region (see text for sequence of primers). (b) Complete nucleotide sequence of candHPV86.

Fig. 1. The genome of candHPV86. (a) The organization of the three cloned overlapping PCR products: plasmid no. 1, plasmid no. 2 and the GP5\(^{\text{&\textregistered}}\)/6\(^{\text{&\textregistered}}\) region (see text for sequence of primers). (b) Complete nucleotide sequence of candHPV86.
GenBank, accession no. AF349909. The nucleotide sequence of the candHPV86 L1 ORF was most closely related to HPVHAN2294 (GenBank, AJ400628; 86% homology) and HPV84 (84% homology), qualifying it as a novel type. The DNA clones and sequence were submitted to the Human Papillomavirus Reference Laboratory (Heidelberg, Germany), and the virus was assigned the number candHPV86. Under the newly proposed terminology, PCR-cloned HPV genomes are considered candidate genomes (personal communication, Dr Ethel-Michele de Villiers).

The candHPV86 genome displayed the same ORF distribution and potential genes found in other sequenced HPV types. The predicted ORFs are summarized in Table 1(a). Table 1(b) shows the homology of putative candHPV86 proteins to the analogous proteins of several closely related HPV types identified by BLAST searches. The candHPV86 proteins were closely related to HPVHAN2294, HPV84, HPV61, HPV72 and HPV83 proteins.

To investigate the relationship between candHPV86 and related HPV genomes, the predicted amino acid sequences of candHPV86 were aligned with the corresponding sequences of HPVs from homology group A3. Sequences were aligned using Sequencer software and verified manually. The resulting phylogenetic tree was calculated and based on the available full-length sequences of HPV genomes from clade A3 (Brown et al., 1999; Chan et al., 1995; Higgins & Sharp, 1988; Terai & Burk, 2001). A representative tree is shown in Fig. 2(a). The tree was consistent with the prior analyses that candHPV86 was most closely related to HPVHAN2294 and placed both candHPV86 and HPVHAN2294 into group A3 with HPV61,
HPV72, HPV83 and HPV84 (Brown et al., 1999; Chan et al., 1995; Terai & Burk, 2001).

The sequence between the end of the L1 ORF and the beginning of the E6 ORF is called the LCR because it contains the origin of replication (ORI) and numerous control signals for DNA replication and transcription. The LCR of candHPV86 is 811 bp in length, whereas the LCRs of HPVHAN2294 and HPV84 are 813 bp and 772 bp, respectively.

Comparison of the LCRs from candHPV86, HPVHAN2294 and HPV84 are shown in Fig. 2(b). Papillomavirus LCRs contain multiple binding sites for transcriptional regulatory factors including AP-1 (Chan et al., 1990), NF-1 (Apt et al., 1993), SP-1 (Gloss & Bernard, 1990), transcriptional enhancer factor (TEF)-1 (Ishiji et al., 1992) and YY-1 (Dong et al., 1994; May et al., 1994). The candHPV86 LCR contains many of these canonical binding sites. The E6/E7 promoter TATA box was identified at positions 7940–7946, 38 bases upstream from the start codon of the E6 ORF.

Papillomavirus ORIs typically contain an E1-binding site between two E2-binding sites (ACCN_4GGT) (Brown et al., 1999; Chow & Leong, 1999; Lu et al., 1993; Sun et al., 1996). The candHPV86 LCR contains three exact E2-binding sites and four E2-binding sites with one base mismatch. A closely related, putative E1-binding site (one base mismatch; candHPV86 and HPVHAN2294 at position 7867 and 7994, respectively) was found within both candHPV86 and HPVHAN2294 LCRs. In addition, there are two TEF-1 sites and a single poly(A) signal. Although the candHPV86 LCR is closely related to the LCR from HPVHAN2294 (73.5%) and HPV84 (73.2%), the organization of many regulatory elements is different, as shown in Fig. 2(b).

candHPV86 was cloned from a clinical specimen by use of the overlapping PCR method (Terai & Burk, 2001). The overlapping PCR technique facilitates obtaining full-length sequences of HPV from samples with low copy numbers. The sequence of candHPV86 was generated using polymerases with low sequence error rates and all predicted ORFs were intact; however, an error rate less than 1 per 1000 cannot be excluded. Novel HPV types are still emerging, and the clinical significance of such types is unknown. The prevalence of candHPV86 and its association with cervicovaginal neoplasia could not be determined since this genome was not included in the MY09/MY11 PCR system (MY09 primer, four bases mismatch at position 7047–7066; MY11 primer, three bases mismatch at position 6615–6634, respectively). The candHPV86-containing phylogenetic branch within the A3 clade includes HPV84 and HPVHAN2294. HPV84 is a highly prevalent type detected in normal and human immunodeficiency virus (HIV)-1 infected women (Terai & Burk, 2001). HPVHAN2294 is similar to HPV partial sequences, uwCerv274-5Sb (98% homology) and HPVXS4 (96% homology), which were detected in the genital tract of an HIV-infected woman and from a hyperkeratotic papilloma on the forearm of a renal transplant recipient, respectively (Berkhout et al., 2000). Based on the phylogenetic tree analysis, it is postulated that candHPV86 is associated with low grade cervical lesions. Detailed characterization of HPV genomes, HPV variants and their organization and comparative analysis of nucleotide and amino acid sequences of viral genes and predicted proteins should provide insight into their biological and clinical behaviour.

This work was supported in part by grants from the NIH to R.D.B. Assignment of the HPV type number was kindly performed by Dr Ethel-Michele de Villiers (Human Papillomavirus Reference Laboratory, Heidelberg, Germany).

References


Received 4 April 2001; Accepted 1 June 2001
Published ahead of print (22 June 2001) in JGV Direct as DOI 10.1099/vir.0.17783-0