



Published in final edited form as:

J Gen Virol. 2007 November ; 88(0 11): 2952–2955. doi:10.1099/vir.0.83178-0.

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

Zigui Chen¹, Mark Schiffman², Rolando Herrero³, and Robert D. Burk^{1,4}

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

Abstract

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 $\alpha 3$ and $\alpha 15$. 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).

Correspondence Robert D. Burk burk@aecom.yu.edu.

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.

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 $\alpha 9$, $\alpha 11$, $\alpha 7$, $\alpha 5$ and $\alpha 6$ 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 $\alpha 3$) and HPV90 (genital HPV species $\alpha 15$), 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 (Forsslund *et al.*, 1999). The 455 bp 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 10^7 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 from closely related representative HPV types, respectively. Sequence alignments were prepared with computer programs CLUSTAL_X software (Thompson *et al.*, 1997) and Codon Align (ver 1.0) (available from website <http://www.sinauer.com/hall/>), as previously described (Chen *et al.*, 2007). For Bayesian tree construction, the gamma model was set for among-site rate variation, and allowed all substitution rates of aligned sequences to be different. Maximum-parsimony (MP) and neighbour-joining (NJ) trees were calculated by a heuristic search in PAUP* v4.0b10

(Swofford, 1998). For maximum-parsimony analyses, amino acid and nucleotide sequences were reduced to phylogenetically informative sites. Data were bootstrap resampled 1000 times. Sequences of related PVs analysed in this work are available from the NCBI database.

A computational algorithm to test the incongruence length difference (ILD) was used to assess the relative degrees of incongruence across different node/partition pairs (Narechania *et al.*, 2005). The partition homogeneity test employed in the PAUP^* v4.0b10 (Swofford, 1998) was performed by treating early and late genes as separate partitions. The ILD values were calculated.

HPV species were generally classified according to de Villiers *et al.* (2004). However, we group HPV71 and HPV90 into the $\alpha 15$ species based on nucleotide similarities as previously described (Narechania *et al.*, 2005; Schiffman *et al.*, 2005).

The two HPV genomes were characterised by overlapping PCR. The complete nucleotide sequences of HPV102 and HPV106 contain 8078 and 8035 bp, with a G+C content of 45.3 and 46.9 mol%, respectively. The complete nucleotide sequences of HPV102 and HPV106 L1 ORFs share 84.1 % and 83.5 % similarity to HPV83 L1 (genital HPV species $\alpha 3$) and HPV90 L1 ($\alpha 15$), respectively. The nucleotide sequence of a small region of the HPV102 L1 ORF shares 99 % similarity with a previously reported MY09/11 PCR fragment, GA6053 (452 bp, GenBank accession no. Y11911) (Astori *et al.*, 1997).

To determine whether the two genomes showed similar degrees of similarity across different regions of their genomes, ratios of relatedness were calculated. The ratio of relatedness of the complete L1:E1 ORFs, using the pairwise similarity of nucleotide sequences of representative HPV types within species $\alpha 3$ and $\alpha 15$, is shown in Table 1. These ratios provide evidence for a similar degree of relatedness across the genomes of species $\alpha 3$ and $\alpha 15$. The L1:E1 ratios of HPV102 and HPV106 compared with related types within the same species are near 1.0 (on average 0.980 and 0.953, respectively), consistent with the L1 ORF being one of the most conserved regions (Terai & Burk, 2002). However, there is a 5' ORF in-frame with the predicted amino terminus of the HPV102 L1 ORF (nt 5353–5784, 432 bp) and the HPV106 L1 ORF (nt 5599–5781, 183 bp) (Supplementary Figure S1, available with the online version of this paper). These sequences overlap the carboxyl terminus of the L2 ORFs and potentially encode amino acids not related to the canonical L1 genes of HPV types within species $\alpha 3$ and $\alpha 15$. Therefore, we do not believe these additional 5' sequences are encoded since there is a consensus translation initiation codon for the L1 ORFs.

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 $\text{CxxC(x)}_{29}\text{CxxC}$, separated by 36 aa. The putative E7 ORFs of HPV102 and HPV106 contain one conserved zinc binding domain, $\text{CxxC(x)}_{29}\text{CxxC}$, 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 (Chan *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 α_2 , α_3 , α_4 and α_{15} were used to calculate Bayesian trees (Fig. 1). HPV32 and HPV42 within the species α_1 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 α_3 and α_{15} , 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 α_3 species forms a sister group to the clade consisting of species α_2 , α_4 and α_{15} , while the 'late ORF' tree indicated the close relationship between species α_3 and α_4 . 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 (α_9) was the most common oncogenic type (3.6 % prevalence), while HPV61 (α_3) and HPV71 (α_{15}) 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 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 HPVs detected in the lower genital tract.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

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

- Apt D, Chong T, Liu Y, Bernard HU. Nuclear factor I and epithelial cell-specific transcription of human papillomavirus type 16. *J Virol*. 1993; 67:4455–4463. [PubMed: 8392590]
- Astori G, Arzese A, Pipan C, de Villiers EM, Botta GA. Characterization of a putative new HPV genomic sequence from a cervical lesion using L1 consensus primers and restriction fragment length polymorphism. *Virus Res*. 1997; 50:57–63. [PubMed: 9255935]
- Castle PE, Schiffman M, Gravitt P, Kendall H, Fishman S, Hildesheim A, Herrero R, Bratti MC, Sherman M. Evaluation of HPV DNA detection by PCR. *J Med Virol*. 2002; 68:417–423. other authors. [PubMed: 12226831]
- Chan WK, Chong T, Bernard HU, Klock G. Transcription of the transforming genes of the oncogenic human papillomavirus-16 is stimulated by tumor promoters through AP1 binding sites. *Nucleic Acids Res*. 1990; 18:763–769. [PubMed: 2156229]
- Chen Z, Schiffman M, Herrero R, Desalle R, Burk RD. Human papillomavirus (HPV) types 101 and 103 isolated from cervicovaginal cells lack an E6 open reading frame (ORF) and are related to gamma-papillomaviruses. *Virology*. 2007; 360:447–453. [PubMed: 17125811]
- Cogliano V, Baan R, Straif K, Grosse Y, Secretan B, El Ghissassi F. Carcinogenicity of human papillomaviruses. *Lancet Oncol*. 2005; 6:204. [PubMed: 15830458]
- de Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen H. Classification of papillomaviruses. *Virology*. 2004; 324:17–27. [PubMed: 15183049]
- Dong XP, Stubenrauch F, Beyer-Finkler E, Pfister H. Prevalence of deletions of YY1-binding sites in episomal HPV 16 DNA from cervical cancers. *Int J Cancer*. 1994; 58:803–808. [PubMed: 7927871]
- Forslund O, Antonsson A, Nordin P, Stenquist B, Hansson BG. A broad range of human papillomavirus types detected with a general PCR method suitable for analysis of cutaneous tumours and normal skin. *J Gen Virol*. 1999; 80:2437–2443. [PubMed: 10501499]
- Gloss B, Bernard HU. The E6/E7 promoter of human papillomavirus type 16 is activated in the absence of E2 proteins by a sequence-aberrant Sp1 distal element. *J Virol*. 1990; 64:5577–5584. [PubMed: 2170687]
- Herrero R, Castle PE, Schiffman M, Bratti MC, Hildesheim A, Morales J, Alfaro M, Sherman ME, Wacholder S. Epidemiologic profile of type-specific human papillomavirus infection and cervical neoplasia in Guanacaste, Costa Rica. *J Infect Dis*. 2005; 191:1796–1807. other authors. [PubMed: 15871111]
- Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*. 2001; 17:754–755. [PubMed: 11524383]
- Ishiji T, Lace MJ, Parkkinen S, Anderson RD, Haugen TH, Cripe TP, Xiao JH, Davidson I, Chambon P, Turek LP. Transcriptional enhancer factor (TEF)-1 and its cell-specific co-activator activate human papillomavirus-16 E6 and E7 oncogene transcription in keratinocytes and cervical carcinoma cells. *EMBO J*. 1992; 11:2271–2281. [PubMed: 1318197]
- Moscicki AB, Schiffman M, Kjaer S, Villa LL. Chapter 5: Updating the natural history of HPV and anogenital cancer. *Vaccine*. 2006; 24(Suppl. 3):S42–S51.
- Narechania A, Chen Z, Desalle R, Burk RD. Phylogenetic incongruence among oncogenic genital alpha human papillomaviruses. *J Virol*. 2005; 79:15503–15510. [PubMed: 16306621]
- Schiffman M, Herrero R, Desalle R, Hildesheim A, Wacholder S, Rodriguez AC, Bratti MC, Sherman ME, Morales J. The carcinogenicity of human papillomavirus types reflects viral evolution. *Virology*. 2005; 337:76–84. other authors. [PubMed: 15914222]
- Swofford, DL. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Sinauer; Sunderland, MA: 1998.
- Terai M, Burk RD. Complete nucleotide sequence and analysis of a novel human papillomavirus (HPV 84) genome cloned by an overlapping PCR method. *Virology*. 2001; 279:109–115. [PubMed: 11145894]
- Terai M, Burk RD. Identification and characterization of 3 novel genital human papillomaviruses by overlapping polymerase chain reaction: candHPV89, candHPV90, and candHPV91. *J Infect Dis*. 2002; 185:1794–1797. [PubMed: 12085327]

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 1997; 25:4876–4882. [PubMed: 9396791]

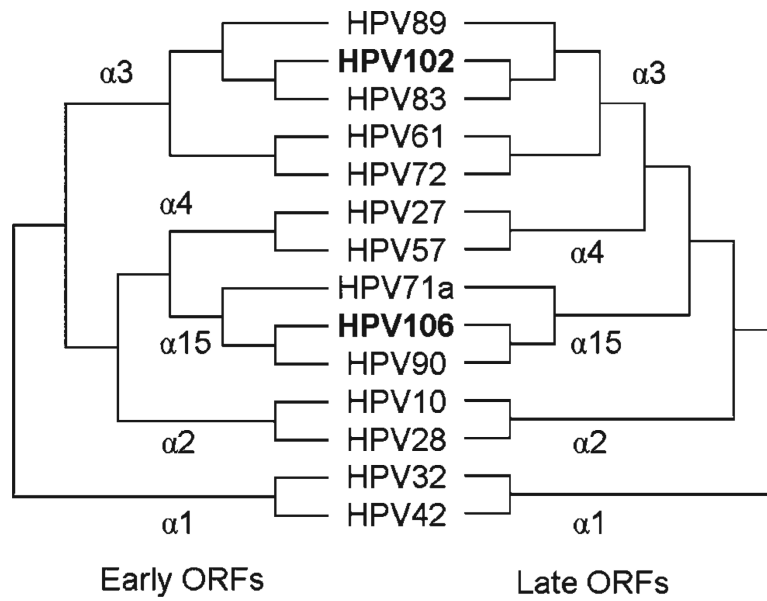


Fig. 1. Bayesian trees inferred from the concatenated amino acid and nucleotide sequences of early (E6, E7, E1, E2 and E4) and late (L2 and L1) ORFs. HPV102 and HPV106, highlighted in bold, are most closely related to HPV83 (84.1 % identity) and HPV90 (83.5 % identity), and are placed in the genital HPV genome homology species $\alpha 3$ and $\alpha 15$, respectively.

Table 1

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

Novel types Related types (species)	Pairwise similarity of each ORF and the URR (%)								L1:E1 ratio
	E6	E7	E1	E2	E4	L2	L1	URR	
HPV102									
HPV83 (α3)	85.0	87.3	85.0	84.9	83.7	78.8	84.1	71.9	0.989
HPV89 (α3)	72.7	72.0	78.3	74.2	73.2	68.4	74.8	64.4	0.955
HPV61 (α3)	66.7	73.4	75.3	73.3	73.9	68.7	75.0	64.7	0.996
HPV72 (α3)	69.8	71.2	76.2	73.8	74.3	69.9	74.7	66.9	0.980
HPV106									
HPV90 (α15)	78.7	85.4	87.3	87.0	85.5	80.4	83.5	73.6	0.956
HPV71 (α15)	68.2	78.7	79.2	76.8	74.3	72.2	75.2	60.8	0.949