Complete nucleotide sequence, genomic organization and phylogenetic analysis of a novel genital human papillomavirus type, HLT7474-S

Vincent T. K. Chow and Peter W. F. Leong

Programme in Infectious Diseases, Department of Microbiology, Faculty of Medicine, National University of Singapore, Kent Ridge, Singapore 117597, Republic of Singapore

A novel human papillomavirus (HPV) type, HLT7474-S, was isolated from a cervical scraping of a female sex worker with a wart virus infection. The complete DNA sequence of 7812 bp was derived from four overlapping PCR products and authenticated by RFLP analysis. The L1 gene exhibited 78% identity to those of its most closely related known HPV types in group A7, comprising HPV types 18, 39, 45, 59, 68 and 70. The genomic organization and phylogenetic analysis of HLT7474-S and group A7 HPVs reiterated their relatedness. Of significance were the strong sequence similarity, phylogenetic relationship and conservation of critical motifs between the transforming E6 and E7 of HLT7474-S and E6 of HPV-18 and E7 of HPV-59, respectively. These features clearly suggest that HLT7474-S is a high-risk genital HPV isolate, closely related to HPV-18 and other members of the A7 group of genital HPVs.

The papillomaviruses are a heterogeneous group of DNA viruses with circular double-stranded DNA genomes of ~8 kb, which infect humans as well as numerous, diverse animal species. To date, about 80 human papillomavirus (HPV) types have been reported, broadly divided into cutaneous and mucosal HPV types. The latter category predominantly infects the genital tract, with genital HPVs further classified as low, moderate or high risk according to their association with genital cancer, especially of the uterine cervix (zur Hausen & de Villiers, 1994; Delius et al., 1998). These sexually transmitted, tumour-virus infections are increasing worldwide, being more prevalent in individuals with impaired cell-mediated immunity; their rising incidence is therefore fuelled in part by the human immunodeficiency virus and AIDS pandemic (Palefsky et al., 1999). Detailed characterization of HPV genomes and their organization and comparative analysis of nucleotide and amino acid sequences of viral genes and proteins can lend some insight into their tissue tropism and risk of carcinogenicity. Here, we describe the isolation, complete nucleotide sequence, genome organization and phylogenetic relationships of a novel genital HPV type from Singapore, the characteristics of which are compatible with those of a high-risk genital HPV.

The source of the viral genome was a cervical scraping taken from a female sex worker undergoing colposcopy in the Department of STD Control Clinic, Singapore, on 23 April 1996. While no malignant cells were observed by cervical cytology, histopathological examination of a concurrent punch biopsy of the cervix demonstrated features of a wart virus infection of the squamous epithelium. Screening of DNA extracted from both cervical specimens (scraping and biopsy) by PCR with consensus primers PCOUP [5' KIKKRACC-GAAACGTT 3', from the long control region (LCR)] and PCVOD (5’ YICIRMAWACTTTCGTTTA 3’, from the E1 gene), designed to amplify DNA of several genital HPVs (including HPV types 6, 11, 16, 18, 31 and 33), produced target fragments of ~1.1 kb. Further, nested PCR with type-specific primer pairs for common genital HPVs types 11, 16 and 18, which have been described previously (Chow et al., 1990a, b; Tham et al., 1991), proved negative. Preliminary DNA sequencing of the ~1.1 kb consensus PCR product by cycle sequencing (Chow et al., 1998) by using 32P-labelled primers and the AmpliCycle Sequencing kit (Perkin Elmer) revealed nucleotide similarities of its 5’ portion to HPV-18 (78% identity) and of its 3’ portion to HPV-59 (77% identity). We then proceeded to obtain the full-length viral genome via a strategy of overlapping PCR fragments amplified from template DNA of the cervical scrape. Thus, by using consensus L1 gene primers MY11 and MY09 (Bauer et al., 1991), purchased from Maxim Biotech, a fragment of ~450 bp was next derived and sequenced. Based on nucleotide sequence information of the two consensus PCR fragments, two pairs of specific primers were synthesized to amplify two overlapping fragments of 1267 and 5796 bp, flanked by primer pairs FC1F/FC1R (spanning nucleotides 6839–6859 and 293–272) and FD1F/FD1R (spanning nucleotides 925–944 and 6720–...
by classical PCR with AmpliTaq DNA polymerase (Perkin Elmer) and by long-distance PCR with Advantage $Th$ polymerase (Clontech), respectively. All four sub-genomic fragments were subjected to cycle sequencing in both directions by using the ABI PRISM BigDye Terminator cycle sequencing ready reaction kit and an ABI PRISM 377 DNA sequencer (Perkin Elmer). To amplify the full-length viral genome from the cervical scrape DNA template by long-distance PCR, a pair of back-to-back L1 gene-specific primers, FE74F (5’ CGGAATTCTCATAAATGTCATCAGTATATA 3’) and FE74R (5’ CGGAATTCTCATTGACAGTATAAAGGTA 3’), were designed to flank a ~7.8 kb target fragment spanning nucleotides 6655–6654. This amplified full-length viral genome was used as template in nested PCR screening experiments using seven pairs of specific primers that targeted overlapping segments, and which also served as sequencing primers. The almost full-length viral DNA was generated by semi-nested PCR with primers NFE74F (nt 6701–6720) and FE74R and the amplified complete viral genomic DNA as template, and then subjected to RFLP analysis with restriction endonucleases Asel, PvuI, SpeI, StuI and TaqI. Computer software for processing sequence data included DNASIS and PROSIS for DNA translation, prediction of open reading frames (ORFs), DNA motif search and hydrophilicity profiles; BCM for mapping of restriction sites; BLAST and GCG for sequence comparison and protein motifs search; CLUSTALW for multiple alignments and PHYLIP for constructing phylogenetic trees.

Fig. 1(a) (lanes 1–4) displays the four overlapping amplified fragments that were sequenced to generate the complete nucleotide sequence of HLT7474-S shown in Fig. 1(b). Although cloning of overlapping PCR products is a feasible way of acquiring full-length HPV sequence data (Forslund & Hansson, 1996), there is a possibility of co-amplifying subgenomic parts of other HPV types present simultaneously in a given cervical specimen. In order to exclude this possibility, the whole viral genome was amplified and confirmed by screening by nested PCR with seven specific primer pairs, which yielded products of expected sizes (data not shown). For further authentication, nearly complete viral genomic DNA of 7766 bp was digested with five informative restriction enzymes and produced RFLP profiles consistent with those predicted from the restriction site mapping derived from the nucleotide sequence (Fig. 1a, lanes 6–10). The viral genome contains 7812 bp with a G+C content of ~38 mol%, well within the expected range for a typical HPV. Table 1 summarizes the locations of the ORFs and their putative encoded proteins, as well as motifs within the LCR, numbered in accordance with the related prototype HPV-18 (Cole & Danos, 1987).

The DNA sequence of the L1 ORF of HLT7474-S shared similarities of 78.3%, 78.1% and 78.0% with those of the most closely related known types, HPV types 39, 70 and 45, respectively, thus satisfying the criteria for a new HPV type, which is defined on the basis of a dissimilarity exceeding 10% in the L1 gene (Delius et al., 1998). In addition, the E6 and E7 ORFs showed highest percentage similarities of 79.7 and 77.9% to E6 of HPV-18 and E7 of HPV-59, respectively, thus reiterating the relatedness between the novel HPV type and known genital HPVs belonging to group A7 (Chan et al., 1995; Cole & Danos, 1987; Volpers & Streec, 1991; Rho et al., 1994; Longuet et al., 1996; Forslund & Hansson, 1996). A 364 bp partial DNA sequence of the L1 gene of HPV strain LIAE5 (GenBank accession no. AF039910) was completely identical to the corresponding nucleotides 6573–6936 of HLT7474-S.

Phylogenetic trees (Ho et al., 1991; Chan et al., 1992) based on the individual ORFs, putative proteins and LCRs were constructed to determine the relationships between HLT7474-S and the known HPVs from group A7. The E1, E2, E5 and L2 proteins of the novel HPV were more closely associated with those of HPV-70 and HPV-39. However, greater similarities were demonstrated for the E4, E6 and L1 proteins of the novel type to those of HPV-18 and HPV-45, and for the novel HPV E7 protein and LCR to its HPV-59 counterparts (Fig. 2).

Alignment of the amino acid residues of the novel HPV putative E6 protein with those of several mucosal and cutaneous HPVs revealed two conserved zinc finger domains (aa 32–68 and 105–141 of E6), each consisting of two CXXC
motifs, which are important for transcriptional activation and transformation by E6. Furthermore, the novel E6 possessed p53-degradation motifs (FAF and RFHKI) and p53-binding regions (RPY and CQKPLCPAEK) that are highly conserved in high- or intermediate-risk genital HPVs, particularly those in group A7 (Crook et al., 1991; Gardiol & Banks, 1998; Rapp &...
Table 1. Features of the ORFs and LCR of the genome of HLT7474-S

<table>
<thead>
<tr>
<th>ORF</th>
<th>Nucleotide sequence</th>
<th>Putative product</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Start</td>
<td>First ATG</td>
</tr>
<tr>
<td>E6</td>
<td>57</td>
<td>105</td>
</tr>
<tr>
<td>E7</td>
<td>579</td>
<td>587</td>
</tr>
<tr>
<td>E1</td>
<td>914</td>
<td>920</td>
</tr>
<tr>
<td>E2</td>
<td>2769</td>
<td>2799</td>
</tr>
<tr>
<td>E4</td>
<td>3412</td>
<td>–</td>
</tr>
<tr>
<td>E5</td>
<td>3925</td>
<td>3940</td>
</tr>
<tr>
<td>L2</td>
<td>4208</td>
<td>4226</td>
</tr>
<tr>
<td>L1</td>
<td>5493</td>
<td>5610</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LCR element or motif</th>
<th>Consensus sequence</th>
<th>Positions of putative sites in LCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>TWNTWATWNHWWWWWAYAAT</td>
<td>7811–18</td>
</tr>
<tr>
<td>E2</td>
<td>ACCGNNNGCGGT</td>
<td>40–51, 56–67, 7456–7467, 7776–7787</td>
</tr>
<tr>
<td>GRE</td>
<td>TACANNNGTGTCT</td>
<td>7800–2</td>
</tr>
<tr>
<td>NF1</td>
<td>TTGC</td>
<td>7530–7534, 7696–7700, 7787–7791</td>
</tr>
<tr>
<td>SPI</td>
<td>NGGNCN</td>
<td>33–38</td>
</tr>
<tr>
<td>TEF1</td>
<td>YRCATDBYDB</td>
<td>7183–7192, 7657–7666</td>
</tr>
<tr>
<td>YY1</td>
<td>MCATNK</td>
<td>7411–7417, 7441–7447, 7658–7664</td>
</tr>
<tr>
<td>TATA box</td>
<td>TATATATAA</td>
<td>71–79</td>
</tr>
<tr>
<td>Poly(A) signal</td>
<td>AATAAA</td>
<td>7249–7254, 7639–7644</td>
</tr>
<tr>
<td>20-mer motif</td>
<td>TGCTTTTAGGCACATATTT</td>
<td>7647–7666</td>
</tr>
</tbody>
</table>

Chen, 1998). It is noteworthy that the latter C-terminal p53-binding motif differed from those of HPV-18 and HPV-45 by only one residue. Interestingly, around nt 234 and 415 of HLT7474-S, we identified E6* splice-donor and -acceptor consensus sites, which are found in certain HPVs associated with anogenital cancers, including HPV types 16, 18, 31 and 33 (Goldsborough et al., 1989). The presence of alternatively spliced E6* mRNA encoding a variant protein E6* may be significant, since it may have different biological properties from the wild-type E6 protein (Rapp & Chen, 1998).

The E7 protein of HLT7474-S harboured three important conserved regions. At aa 42–108 there was a metal-binding domain that has been shown to be important for the transforming ability, dimerization and stability of the E7 protein. The metal-binding domain contained two CXC zinc finger motifs, the C-terminal motif of HPV-16 E7 having been shown to be essential for its transactivating function as well as for the immortalization of human keratinocytes. The pRb-binding and CKII phosphorylation motifs were detected within the CR2 domain, which is involved in the efficient dissociation of E2F from pRb as well as in the induction of DNA synthesis.

The occurrence of aspartic acid as the first residue of the pRb-binding motif (DLYCYEE) is notable, given that this residue occurs in high-risk HPVs and is critical for the transforming activity of E7. The serine residue in the CKII phosphorylation site (NSEEEIDE) has also been implicated in transformation by E7. The CR1 domain (MHGPKPTVHIVLDL) of the novel E7 protein shared several conserved residues found in genital HPVs, especially HPV-18, the sequence of which differed only by four of 15 residues (Munger & Halpern, 1997).

The motif MXYXH, which is conserved in genital HPVs (Chan et al., 1995), was detected in the L1 protein of the novel HPV. Similarly, the L2 protein possessed the motif TTPA(I/V)L(D/N)(I/V), which is highly conserved in mucosal HPVs and is thought to play a role in tissue tropism (Rho et al., 1994). Comparison of the hydrophilicity plots of the L1 proteins of HLT7474-S and other HPVs within group A7 showed strikingly similar profiles despite some differences in amino acid sequence. The corresponding L2 hydrophilicity plots depicted generally similar patterns, albeit with distinct variations, especially in the C-terminal third of L2. These observations are congruent with the notions that the L1 major
A novel genital HPV type, HLT7474-S

Fig. 2. Phylogenetic trees based on sequences of HPV proteins and LCRs to illustrate and compare the relatedness of HLT7474-S to HPV types belonging to group A7.

capsid proteins constitute group-specific antigens of papillomavirus while the L2 proteins may act as type-specific antigens.

Spanning nt 7125–104, the HLT7474-S LCR contains a diverse array of signature motifs, including the TATA box representing the E6/E7 promoter, the polyadenylation signal and putative binding sites for E1, E2, transcription factors AP1, NF1, SP1, TEF1 and YY1, as listed in Table 1. The putative origin of replication, to which the E1 protein binds to initiate DNA replication, was identified within the 3′ segment of the LCR upstream of the TATA box. There were four binding sites for the E2 protein, which regulates viral DNA replication and gene expression. Of additional interest is a sequence spanning nt 7800–2, conforming to the consensus motif TACANNN-TGTCTT of the glucocorticoid-responsive element (GRE), which may serve to upregulate expression from the E6/E7 promoter (O’Connor et al., 1995). Also of note within the LCR is the existence of a 20-mer motif TGCTTTTAGGCACATT (nt 7647–7666), which as yet has no characterized function but which is extremely well-conserved in intermediate- and high-risk anogenital HPV types and is proposed to contribute to their oncogenicity (Marich et al., 1992).

In conclusion, we have isolated a novel genital HPV type, HLT7474-S, and characterized its complete genomic sequence.
and organization and its putative proteins. The features of this isolate warrant its designation as a co-evolved member of the A7 group of genital HPV's. Taken together, the phylogenetic relationships and the occurrence of conserved motifs within the individual viral genes, their encoded proteins and the LCR suggest strongly that this novel HPV type is potentially a highly oncogenic virus with a predilection for genital mucosa. Of conspicuous prominence is the remarkable resemblance of the sequences and conserved motifs of its transforming E6 and E7 proteins with those of HPV-18, which has a well-established causative link with highly malignant genital cancers associated with poor prognosis (Burger et al., 1990; Comerford et al., 1995). Finally, the availability of sequence information for this novel HPV may facilitate future experiments in vitro as well as clinical studies to ascertain its prevalence of infection and carcinogenic potential (Meyer et al., 1998).

We are grateful to Dr Roy Chan (National Skin Centre, Singapore) for providing the cervical specimens and to W.M. Yeo and S.Y. Tan for technical assistance. This work was supported by a grant from the National Medical Research Council, Singapore (NMRC/0163/1996).

References


A novel genital HPV type, HLT7474-S


Received 2 June 1999; Accepted 3 August 1999