Journal of General Virology (1999), **80**, 2923–2929. Printed in Great Britain SHORT COMMUNICATION

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.

Author for correspondence: Vincent Chow.
Fax +65 776 6872. e-mail micctk@nus.edu.sg

The GenBank accession number of the sequence reported in this paper is AF131950.

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 PVCOU [5′ KKIKKRACC-GAAAICGGT 3′, from the long control region (LCR)] and PVCOD (5′ YIICIRMAWACTTTCGTTTTA 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 HPV types 11, 16 and 18, which have been described previously (Chow *et al.*, 1990*a*, *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 ^{33}P-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–

6701), by classical PCR with AmpliTaq DNA polymerase (Perkin Elmer) and by long-distance PCR with Advantage *Tth* 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′ CGGAATTCCTAATCCAGTTCCATCTATA 3′) and FE74R (5′ CGGAATTCTAGTTGCAGTAGATAAGG-TA 3′), were designed to flank a $\sim$ 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 *Ase*I, *Pvu*I, *Spe*I, *Stu*I and *Taq*I. 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. 1*a*, lanes 6–10). The viral genome contains 7812 bp with a G+C content of $\sim$ 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 & Streeck, 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 L1AE5 (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

(*a*)

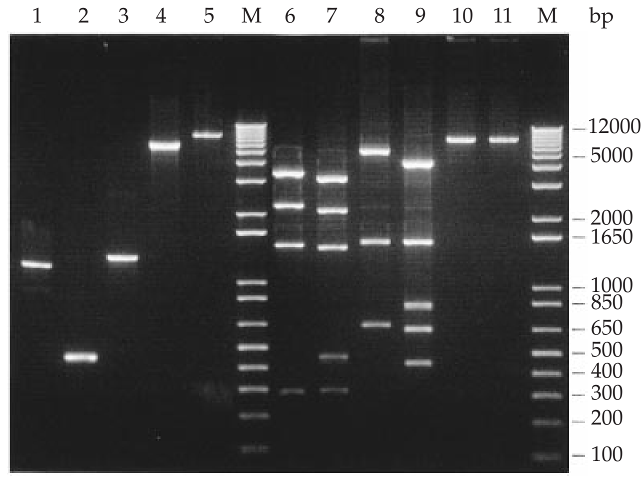

Fig. 1. PCR amplification, RFLP analysis and genomic sequence of HPV DNA of HLT7474-S. (*a*) Agarose gel electrophoresis of overlapping PCR products of $\sim$ 1·1 kb, $\sim$ 450 bp, 1267 bp and 5796 bp (lanes 1–4) and of the amplified full-length genome of $\sim$ 7·8 kb (lane 5). Restriction fragments of expected sizes following complete digestion of the almost full-length viral genomic DNA (spanning nt 6701–6654) with *Ase*I (315, 1520, 2334, 3597 bp), *Spe*I (321, 478, 1488, 2188, 3291 bp), *Stu*I (685, 1577, 5504 bp) and *Taq*I (447, 651, 835, 1570, 4263 bp) are depicted in lanes 6–9, respectively. The 7766 bp template remains uncleaved after *Pvu*I digestion, confirming the absence of *Pvu*I sites in the genome (lane 10). Lanes 11 and M contain undigested template DNA and the 1 kb PLUS DNA ladder marker (Gibco BRL). (*b*) Complete nucleotide sequence of HPV genome of HLT7474-S. Nucleotide 1 is based on alignment with HPV-18.

(*b*)

```
   1 CTTATACTTT TTTATAATTT TATTACTGAA AAGGGTGTAA CCGAAAACGG TCATGACCGA AATCGGTGCA TATATATAAC CCAGCTTACA GTACATGTAA
 101 TAGGATGGCT GAATTTGGCA ATCCAGCTAC CCGACCCTAC AAACTACCAG ACCTATGCAA CACACTGGAC ACATCACTGC AAGATATAGA AATAAGCTGT
 201 GTATATTGTA AAAGCGTCCT GCAGCGAACA GAGGTATATG AATTTGCCTT TGCTGATTTA TTTGTAGTAT ATAGAGACGG TATACCATAT GCAGCATGTC
 301 AGAACTGTTT AATGTTTTAT TCTAAAATAA GAGAACTAAG ATATTATAGC GACTCTGTGT ATGGGGAAAC GTTAGAAAAA CTAACAAATA GCAATATATA
 401 TGATATATTA ATAAGGTGTT TACGGTGTCA AAAACCTTTA TGTCCGGCAG AAAAACTAAA ACACCTTAAT GAAAAACGAA GATTCCATAA AATAGCAGGG
 501 AAATATAGAG GACAGTGTCG CCGGTGCATG ACGCGTGCAC AAGAACAGCA AGGCAGCCGC AGAGAAACAC AAGTATAAGG GTCGCTATGC ATGGACCAAA
 601 ACCTACTGTA CATGACATTG TGTTAGATTT GGAACCATAT AATGAAGTGC AAGAGGTTGA CCTGTATTGC TATGAGGAAT TAAACAACTC AGAGGAAGAA
 701 ATAGATGAAC CAGATAATGC AATTAATCAC CGACAACCAC TACTAGCCAG ACGAGAAGAA CTACAGCGTC ACACTATTTG CTGCGTGTGT TGTAAATGTG
 801 AGGCCAGTCT GCAGTTAGTG GTGGAAAGTT CAGCAGCCGA TCTACGAGAT CTCCAGCAGC TGTTTTTGGG CACGCTATCA TTTTTGTGTC CACTGTGTGC
 901 AGTACTGCGG TAAGCTGCAA TGGCCGACCC TGAAGGTACA GATGGGGATG GGACGGGATG TAATGGTTGG TTTTTTGTGC AGGCAATAGT AGACAAAAAA
1001 ACAGGTGACA CCATATCAGA GGATGAGGAC GAGGACGGGA CAGATACAGG TTCAGACTTG GTAGATTTTA TTGACACTAG TAATACAAAT TATATGCAGG
1101 CAGATAGAGA GGCAGCACAA GCGCTATTAC ATGCACAGGA AGTAGAAACT GATACAAAGC TATTACATGC CTTAAAACGA AAGTATGGAG CACACAGCAC
1201 AGAAAACAGT CCATGTCGTG ATACAGCAAG TATACATAGC AATTTAAGCT CACCATTACA AGAAATATCG TTAAATAGCT ATAATAATAC AGCTAAACGA
1301 AGGCTGTGTT CAGTGCCAGA CAGCGGCTAT GGCAATACTC AAGTGGAAAC TCTGCAGACT CAGGTAACCC TAGATACAAA TGTGTTTGGG GATGGGAAAA
1401 ATGGCGACGG TTTAAATAGT GAGGCATGTA GTACAGATAA TGAAATGGAT ATAGAAAATC AAAATCCAAA CTCACCAATG ACACAAATTG TGTCCTTATT
1501 AAAAGTAAAT AATAAAAAAG CAGCAATATT GGCTAAATTT AAAGAAAACAT ATGGATTGTC ATTTACAGAT TTGGTACGAA CATTTAAAAG TGATAAAACA
1601 ACATGTACAG ATTGGGTAGC CGCTATTTGT GGAGTAAATC CAAATATTGC AGAAGGATTT AAAACGTTAA TACAACCATA TGTGTTATAT GCACATATAC
1701 AATGCATGGA TTGTTCATGG GGAGTATTTA TATTAGCTTT ATTAAGATAT AAATGTGGCA AAAATAGACT AACTGTAGCT AAAGGACTAA GTACATTACT
1801 ACATGTACCA GATACCCATA TGTTAATTGA GCCACCAAAA CTGCGTAGCA GTTGTGCAGC ACTATATTGG TATAGGACAG GAATATCAAA TATTAGTGAA
1901 GTAACAGGAG ACACACCTGA GTGGATACAA AGACAAACTA TTATACAACA TGGTATAGAT GACAGTGTAT TTGACCTGTC AGAAATGATA CAGTGGGCGT
2001 TTGATAATGA CTATATAGAT GAAAGTGACA TAGCATATGA ATATGCACAG TTAGCTGATT GTAATAGTAA TGCAGCAGCA TTTTTAAAAA GTAATTGTCA
2101 AGCTAAATAC TTACGAGATT GTGCAGTTAT GTGTAGACAT TATAAAAGGG CACAACGAAA GCAAATGAAT ATGTCACAAT GGATTAGTTA TAGATGTGAC
2201 AAAATTGATG ATGGAGGTGA TTGGAAACCT ATTGTACAAT TTTTAAGATT TCAAGGAATA GAGTTTATTA CATTTTTAAG AGCATTTAAA GACTTTTTAA
2301 AAGGAACACC AAAAAAAAAT TGTATTGTTA TATATGGACC AGCAAATACA GGCAAATCAT ATTTTTGTAT GAGTCTTATA CAATTTTTAC ATGGTACAGT
2401 GTTGTCATTT GTAAATTCAA ACAGCCATTT TTGGCTGGAA CCGTTAACAG ACACTAAGAT AGCTATGGTA GATGATGCAA CACCAACATG CTGGTCCTAT
2501 TTTGATAATT ATATGAGAAA TGCATTAGAT GGCAATCCAA TAAGTATAGA TAGAAAACAT AAACATTTAA TACAAATGAA GTGTCCACCA ATGCTAATTA
2601 CATCTAATAC AAATCCTGCA ACAGATGATA GATGGCCTTA TTTACGAAGT AGGGTAACAG TATTTACATT TCCACATACA TTTCCATTTG ATAGCAATGG
2701 CAATCCAGTG TATGACATAA ATGATAAAAA TTGGAAATGT TTTTTTAAAA GGACGTGGTC CAGATTAGAT TTGCACCAGG AAGAGGAGGA TACAGAAAAT
2801 GATGGATGCC CTATCCCAAC GTTTAAATGT GTTACAGGAA AAGATTCTAG AACATTACGA GACTGATAGT ACAGACATAT GTGATCAAAT AGACTATTGG
2901 AAGTGTGTAC GTCTGGAAAA TGCAATATAT TATGCAGCAA GGGAGCATGG TCTTAAAACA ATAAACCACC ACGTGGTACC AACATTTCAA ATTTCAAAAA
3001 GCAAAGCACA TGAAGCAATT GAACTGCAGA TGGCACTAGA GAGTCTTGCT AAATCAGAGT TTAAAAATGA ACTGTGGACA TTGCAAGATA CGTGCCAAGA
3101 ACTATATCAA ACACCTCCAC AACAGTGTTT TAAAAAACAG GGACAAACCG TCGAAGTACG GTATGATGGA GACAAAGACA ATACTATGCA TTATACATCT
3201 TGGGACTATA TATATTATGT AACAGAAGGT GATAAGTGGT GTAAAACCAA AGGTTATGTG AATTATTGTG GATTGTATTA TATAAAGGAG GGACAACAAA
3301 CATATTATGT ACAGTTTAAA TGTGATGCAC AACAATATGG ACAGAGCGGT AAATGGGAAG TGTGGTATAA TGGTAAAAAA ATTGAATGTG TAAATACCTG
3401 TGACAAAATA ATACAATGTT CTGAATCTAT GTACAGTACC TGTGACGAGA CAGTATCCGC TACTGCAATT GCTAGAGAAT TACAACACCC CACCACACCG
3501 TATACCGAAG CCACCACCGT GTGCACCCAA AAGAGCGGGG GTTCGGCGCC GACTAGAAAC CCATTCAGAC ATTGTGGATT CACAGAAACA AGTGAAGTCG
3601 ACGGACTGTC CGTGGACCAC CTCAACAACC CCCTGCTTAG TGCACCTGCA GGCAACAACG GGTTCAGGAA GCACAATAGT GGTGACACTA CGCCTATAGT
3701 GCACCTAAAA GGTGATAAAA ATAGATTAAA GTGTTTGCGA TACAGGCTAC AAAAGTTCCA TAATTTATAT ACACAAATAT CATGTACGTG GCATTGGATA
3801 GGTGGTAAAG AAACAACAAG GACTGGTATA TTAACATTGG CATATGCAAA TGAAACCCAA AGACAAAAAT TTTTGGATGT TGTAAAAATA CCTAATACTG
3901 TTAATGTATC CCTTGGTTAT ATGACATTGT AATAGGGGTA TGTTAATTGC TGTATTTTTG TTGTGCTTTT GTGTGTGCCT ATATGTGTGT TGCACTATTC
4001 CGCTTTTTCC TGCTACTAAT GTGTGTGCGC ATGCTTGGCT ACTTGTGTTT TTGTTAATAG TTTTTATTAC TACCCCTTTT ACTGCATTTA TTGTATATAT
4101 ATTTTTTTTT ATATTGCCAA TGTTTTTATT GCATTTGCAC AGTGTACATA CATTTGACTA AGTACTATAT AACCTGTATT GTACAGGACA TTACATTGGT
4201 AGTGTAATAC CTTTTTGTAA TAAACATGGT ATCCCATCGT GCTGCCCGTC GAAAGCGTGC ATCTGCAACG GATTTGTATA AAACATGTAA ACAATCGGGT
4301 ACATGCCCTC CAGATGTTAT TAATAAGGTT GAAGGTACTA CTCTTGCAGA TAAACTATTA CAATGGACTA GTTTGGGAAT ATTTTTAGGA GGCCTTGGAA
4401 TAGGTACTGG TAGTGGCACA GGGGGGCGTA CCGGCTATAT TCCATTGGGT GGCAGACCTA ATACTGTTGT TGATGTGTCT CCTGCAAGGC CACCTGTTGT
4501 AATTGAATCT GTGGGGCCTT CAGATCCATC TATTGTTACA TTAGTTGAAG AGTCCAGTAT TGTTACATCT GGAGCTCCTG TTCCTACATT TACAGGCACA
4601 TCTGGATTTG AAATTACATC CTCTGCAACA ACCACCCCTG CAGTATTAGA CATTACACCT GCTTCTGGGT CTGTGCAACT AAGTAGCACT AGTTTTACCA
4701 ATCCTGCATT TACCGACCCT TCGGTTATTG AGGTTCCTCA AACAGGGGAG GTGTCTGGAG ATATTTTTAT TACTACCCCT ACATCTGGGA CACATGGATA
4801 CGAGGAAATT CCAATGCACA CATTTGCAAC ACAGGGCAGA GGCACCGAAC CTATTAGTAG TACCCCTATT CCTGGTGTTA GGCGTGTGGC AGGACCTAGA
4901 TTATATAGTC AAGCGTATCA ACAAGTTAAA ATAACTAATT CAGACTTTAT ATCCCGTCCA TCTACGTTGG TTACATTTAC CAATCCTGCA TATGAGCCTA
5001 TAGACACTAC ATTAACCTTT TCACCACAGG ATGTTGTGCC TGATCCTGAT TTTATGGATA TTGTTCGTTT ACATAGGCCT GCCTTAACAT CCAGACGTGG
5101 TACAGTTAGA TTTAGTAGAT TGGGTAAAAA ACTAACTATG TCTACTCGCA GTGGTAAACA AATAGGTGCC CAGGTGCATT ATTATCATGA TATTAGTCCT
5201 ATTTCACATA TAGGTGAAAG TATTGAAATG CAGCCTTTAC TGCCTGATGC AGCAGTTACT GCTGACACTA ATGGCCTCTT TGACATTTAT GCTGACACTG
5301 ATATTGATAA TAATGCCATG CTATATGATA GAAATATTTC TGATGTTACA CAACCTACCA CTTCAACTAT ATCTAGTGTA TCCTCTCGTT ATAGTAATAC
5401 TACTATTCCT TTAGCAACAT CTTGGGATGT TCCTGTTCAT ACAGGGCCTG ACATGACATT ACCTACTACT ATACCCCAGT GGCCTAATAT AGTACCTTTA
5501 CTGCCTAATA ATACACATTC AGTTGTACTT CAGGGAACAA ACTATTATTT ATGGCCTAAT TATTATTTTA TTTTCAAAAA ACGTAAACGT GTTCCCTATT
5601 TTCTTACAGA TGGCTTTGTG GCGTTCTAGT GACAGCAAGG TATACCTTCC ACCACCTTCA GTGGCTAAAG TTGTCAACAC AGACGATTAT GTAACACGTA
5701 CCAGTACATT TTATCATGCT GGCAGCTCTA GGCTTCTAAC CGTTGGACAT CCATACTATA AAGTTACCTC AAATGGAGGC CGCAAGCAAG ACATTCCTAA
5801 AGTGTCTGCC TATCAGTATC GAGTGTTTCG GGTTACATTA CCTGACCCTA ATAAATTTGG TCTTCCAGAC ACTAATATAT ATAATCCTGA AACACAACGT
5901 TTAGTATGGG CCTGTGTTGG CATGGAGGTA GGTCGTGGGC AGCCTTTAGG TGTTAGCCTT AGTGGTCATC CCTTTTACAA CAAATTGGAT GACACAGAAA
6001 ATTCCCATGT TGCTACTTCT GTAGTTACAC ACGACACTAG AGATAATGTG TCAGTGGATT ATAAACAAAC CCAGTTATGT ATTATTGGTT GTGTTCCTGC
6101 TATTGGGGAA CATTGGGCTA AGGGTACTGC CTGTAAGCCC GGTGCTGTGC AAACAGGTGA CTGTCCTCCA TTAGAACTAG TAAATACACC TATTGAGGAT
6201 GGTGATATGA TTGATACTGG CTATGGTGCT ATGGACTTTA GTACTTTGCA GGATAATAAA AGTGAGGTTC CCTTAGACAT TTGTCAATCT ATTTGTAAAT
6301 ATCCAGACTA TTTGCAAATG TCTGCAGATG CATATGGGGA CAGTATGTTT TTTTGTTTAC GACGCGAACA GTTATTTGCA CGCCATTTTT GGAACAGGGG
6401 AGGCACTATA GGAGATGCAG TACCGGAAAC ATTATATATA AAAGGTACTA ATGATAGGGC AACACCTGGA AGCTGTATTT ATTCTCCATC ACCTAGTGGG
6501 TCTATGGTAT CTTCAGATGC ACAAATGTTT AATAAGCCTT ATTGGCTGCA CAAAGCCCAG GGACAAAATA ATGGCATTTG CTGGCACAAC CAATTATTTA
6601 TAACTGTGGT AGACACAACA CGTAGTACCA ATCTTACCTT ATCTACTGCA ACTACTAATC CAGTTCCATC TATATATGAA CCTTCTAAAT TTAAGGAATA
6701 CACACGCCAT GTAGAGGAAT ATGATTTACA ATTTATATTT CAATTGTGTA AAATTACACT TACTACTGAT GTTATGTCTT ATATACATAA CATGGATCCT
6801 ACTATTTTAG ATAGTTGGAA TTTTGGTGTT AGTCCTCCCC CATCTGCTAG CTTAGTAGAT ACATATAGGT TTTTACAGTC ATCTGCCATT ACATGTCAGA
6901 AGGATGTGGT TGTTCCACAA AAAAAGGATC CATATGAAAA ATTAAAGTTT TGGAATGTGG ATCTTAAAGA ACATTTTTCA TCTGATTTAG ATCAGTTTCC
7001 TTTAGGACGT AAGTTTTTAT TACAGGCTGG GTTACGACCT AAACCCACCA TAGGCCCTAG GAAACGTGTT GCTTCTACTT CTACTGCTAC TAGGCCCTCC
7101 AAACGGAAAC GTACTGCTAA ATAATGTGTT GTGTATAATT GTGTTGTTTG TTTTGTTATA TGTGTGTTCC TTATATGTTG TGTGCATGTT TGTATGGTTA
7201 TGTGTATGTT TGTATCATGT GTAAGTAATG TGTGTATTGT ATGTGTTTAA TAAAGTATGT GTGTTAGTTT CGTGTGTGGT TGCACCCAAA TGAGTAAGAG
7301 ACTGTCCCTT TATTATCCCT TGTATCCTGT GTATCCTGTG TGCCCTTTGT TCCTACTTTA TATAGGGTGT GTTATATGTT ATACATTATT TATAATACAC
7401 TTTGTAGCGT CCATTTTATC CATTTGTATC CTACAAGCCT CCATTTTACT ATGCAACCGA TATCGGTTGC CTTTGACACA CCTAACATTA TACATTCTTA
7501 TTGTTACATT ATTTAAAGTT AAATCCTTTT TGGCTCTGTT TAAGGCGCCA GTGAGCCAGT GGCGCACCTT ATATTACTCA TCATCCTGTC CAGGTGCACT
7601 GCAACAATGG TTTGCAAAAC CTTGTTTTAC CCTTACATAA TAAAACTGCT TTTAGGCACA TATTTTCACT GTTTTTACTT GCTTTAATTA CACTATTGGC
7701 CTGTACAACT ACTTTTAGAT TCAAGAATGT GTCTTGTAGG TTATATACCT GTCACTGATT CATACTTTTA TTGCAACCGA TTTCGGTTGG CCAAATCCTT
7801 ACACATTATA TA
```

Fig. 1. For legend see facing page.

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

ORF	Nucleotide sequence				Putative product	
	Start	First ATG	End	Size (nt)	Amino acids	Molecular mass (kDa)
E6	57	105	578	474	157	18·4
E7	579	587	913	327	108	12·4
E1	914	920	2866	1947	648	73·3
E2	2769	2799	3932	1134	377	43·4
E4	3412	–	3699	288	95	10·6
E5	3925	3940	4161	222	73	8·4
L2	4208	4226	5629	1404	467	50·4
L1	5493	5610	7124	1515	504	56·4

LCR element or motif	Consensus sequence	Positions of putative sites in LCR
AP1	TKWNTMA	11–17, 7132–7138, 7211–7217, 7218–7224, 7244–7250, 7247–7253, 7291–7297, 7389–7395, 7537–7543, 7574–7580, 7756–7762
E1	TWNTWATWNHWWWYWAYAAT	7811–18
E2	ACCGNNNNCGGT	40–51, 56–67, 7456–7467, 7776–7787
GRE	TACANNNTGTTCT	7800–2
NF1	TTGGC	7530–7534, 7696–7700, 7787–7791
SP1	NGGNGN	33–38
TEF1	YRCATDBYDB	7183–7192, 7657–7666
YY1	MCATNKT	7411–7417, 7441–7447, 7658–7664
TATA box	TATATATAA	71–79
Poly(A) signal	AATAAA	7249–7254, 7639–7644
20-mer motif	TGCTTTTAGGCACATATTTT	7647–7666

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 CXXC 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 (MHGPKPTVHDIVLDL) 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

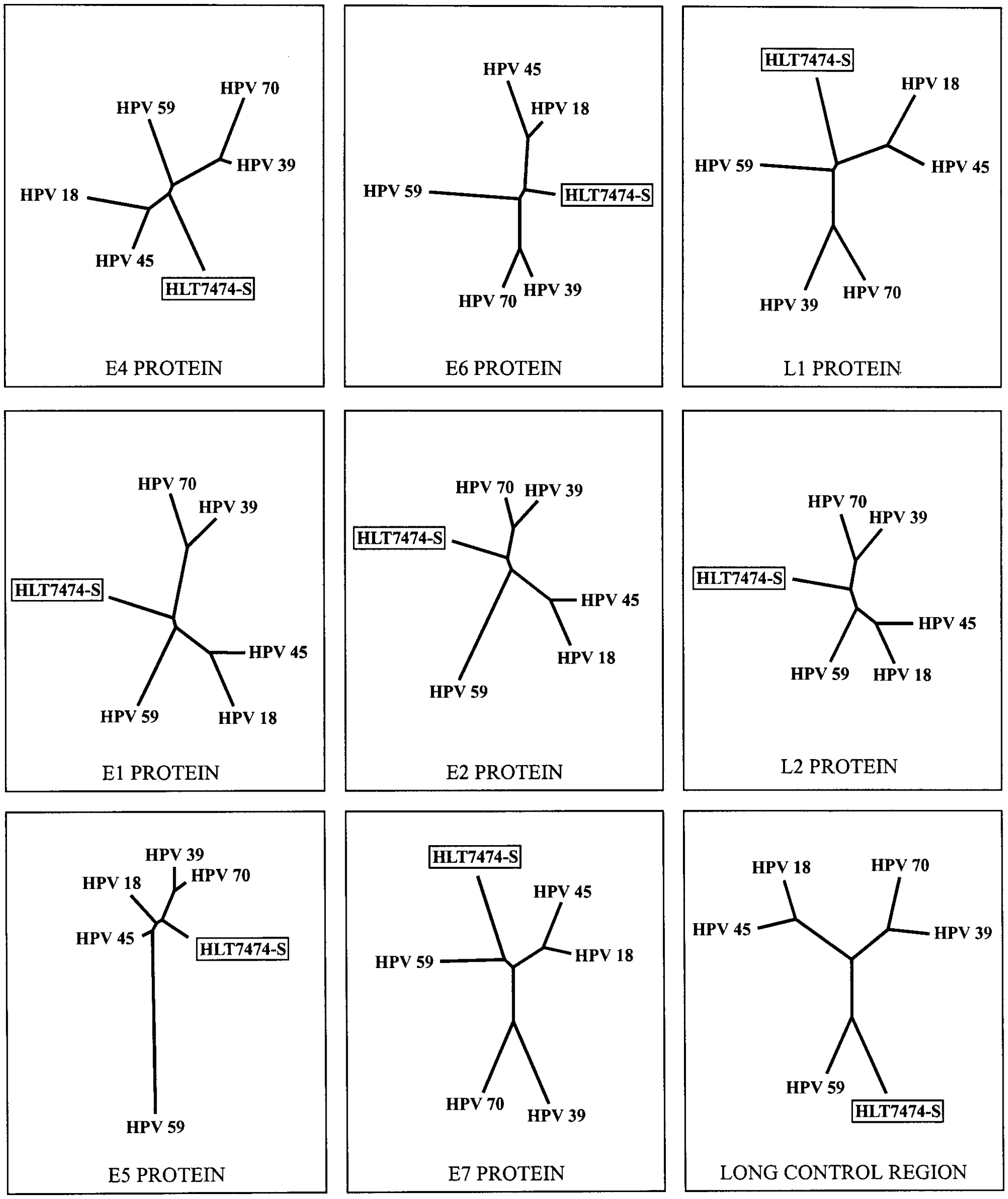

Fig. 2. Phylogenetic trees based on sequences of HPV proteins and LCRs to illustrate and compare the relatedness of HLT7474-S to HPV types belonging to group A7.

capsid proteins constitute group-specific antigens of papillomaviruses 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-TGTTCT 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 TGCTTTTAGGCACAT-ATTTT (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 HPVs. 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*., 1996; 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

Bauer, H. M., Ting, Y., Greer, C. E., Chambers, J. C., Tashiro, C. J., Chimera, J., Reingold, A. & Manos, M. M. (1991). Genital human papillomavirus infection in female university students as determined by a PCR-based method. *Journal of the American Medical Association* **265**, 472–477.

Burger, R. A., Monk, B. J., Kurosaki, T., Anton-Culver, H., Vasilev, S. A., Berman, M. L. & Wilczynski, S. P. (1996). Human papillomavirus type 18: association with poor prognosis in early stage cervical cancer. *Journal of the National Cancer Institute* **88**, 1361–1368.

Chan, S. Y., Ho, L., Ong, C. K., Chow, V., Drescher, B., Durst, M., ter Meulen, J., Villa, L., Luande, J., Mgaya, H. N. & Bernard, H. U. (1992). Molecular variants of human papillomavirus type 16 from four continents suggest ancient pandemic spread of the virus and its coevolution with humankind. *Journal of Virology* **66**, 2057–2066.

Chan, S. Y., Delius, H., Halpern, A. L. & Bernard, H. U. (1995). Analysis of genomic sequences of 95 papillomavirus types: uniting typing, phylogeny, and taxonomy. *Journal of Virology* **69**, 3074–3083.

Chow, V. T. K., Tham, K. M. & Bernard, H. U. (1990*a*). *Thermus aquaticus* DNA polymerase-catalysed chain reaction for the detection of human papillomaviruses. *Journal of Virological Methods* **27**, 101–112.

Chow, V., Tham, K. M., Yeo-Gloss, M., Lim-Tan, S. K., Sng, I., Thirumoorthy, T. & Bernard, H. U. (1990*b*). Molecular diagnosis of genital HPV DNA types by polymerase chain reaction and sensitivity-standardized filter *in situ* hybridization in randomly sampled cohorts of Singapore women. *Molecular and Cellular Probes* **4**, 121–131.

Chow, V. T. K., Chan, Y. C., Yong, R., Lee, K. M., Lim, L. K., Chung, Y. K., Lam-Phua, S. G. & Tan, B. T. (1998). Monitoring of dengue viruses in field-caught *Aedes aegypti* and *Aedes albopictus* mosquitoes by a type-specific polymerase chain reaction and cycle sequencing. *American Journal of Tropical Medicine and Hygiene* **58**, 578–586.

Cole, S. T. & Danos, O. (1987). Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome. Phylogeny of papillomaviruses and repeated structure of the E6 and E7 gene products. *Journal of Molecular Biology* **193**, 599–608.

Comerford, S. A., Maika, S. D., Laimins, L. A., Messing, A., Elsasser, H. P. & Hammer, R. E. (1995). E6 and E7 expression from the HPV 18 LCR: development of genital hyperplasia and neoplasia in transgenic mice. *Oncogene* **10**, 587–597.

Crook, T., Tidy, J. A. & Vousden, K. H. (1991). Degradation of p53 can be targeted by HPV E6 sequences distinct from those required for p53 binding and trans-activation. *Cell* **67**, 547–556.

Delius, H., Saegling, B., Bergmann, K., Shamanin, V. & de Villiers, E. M. (1998). The genomes of three of four novel HPV types, defined by differences of their L1 genes, show high conservation of the E7 gene and the URR. *Virology* **240**, 359–365.

Forslund, O. & Hansson, B. G. (1996). Human papillomavirus type 70 genome cloned from overlapping PCR products: complete nucleotide sequence and genomic organization. *Journal of Clinical Microbiology* **34**, 802–809.

Gardiol, D. & Banks, L. (1998). Comparison of human papillomavirus type 18 (HPV-18) E6-mediated degradation of p53 *in vitro* and *in vivo* reveals significant differences based on p53 structure and cell type but little difference with respect to mutants of HPV-18 E6. *Journal of General Virology* **79**, 1963–1970.

Goldsborough, M. D., DiSilvestre, D., Temple, G. F. & Lorincz, A. T. (1989). Nucleotide sequence of human papillomavirus type 31: a cervical neoplasia-associated virus. *Virology* **171**, 306–311.

Ho, L., Chan, S. Y., Chow, V., Chong, T., Tay, S. K., Villa, L. L. & Bernard, H. U. (1991). Sequence variants of human papillomavirus type 16 in clinical samples permit verification and extension of epidemiological studies and construction of a phylogenetic tree. *Journal of Clinical Microbiology* **29**, 1765–1772.

Longuet, M., Beaudenon, S. & Orth, G. (1996). Two novel genital human papillomavirus (HPV) types, HPV68 and HPV70, related to the potentially oncogenic HPV39. *Journal of Clinical Microbiology* **34**, 738–744.

Marich, J. E., Pontsler, A. V., Rice, S. M., McGraw, K. A. & Dubensky, T. W. (1992). The phylogenetic relationship and complete nucleotide sequence of human papillomavirus type 35. *Virology* **186**, 770–776.

Meyer, T., Arndt, R., Christophers, E., Beckmann, E. R., Schroder, S., Gissmann, L. & Stockfleth, E. (1998). Association of rare human papillomavirus types with genital premalignant and malignant lesions. *Journal of Infectious Diseases* **178**, 252–255.

Munger, K. & Halpern, A. (1997). HPV16 E7: primary structure and biological properties. In *Human Papillomaviruses 1997*, pp. III 17–36. Edited by G. Myers, F. Sverdrup, C. Baker, A. McBride, K. Munger & H. U. Bernard. Los Alamos, NM: Los Alamos National Laboratory.

O'Connor, M., Chan, S. Y. & Bernard, H. U. (1995). Transcription factor binding sites in the long control region of genital HPVs. In *Human Papillomaviruses 1995*, pp. III 21–40. Edited by G. Myers, H. U. Bernard, H. Delius, C. Baker, J. Icenogle & A. Halpern. Los Alamos, NM: Los Alamos National Laboratory.

Palefsky, J. M., Minkoff, H., Kalish, L. A., Levine, A., Sacks, H. S., Garcia, P., Young, M., Melnick, S., Miotti, P. & Burk, R. (1999). Cervicovaginal human papillomavirus infection in human immunodeficiency virus-1 (HIV)-positive and high-risk HIV-negative women. *Journal of the National Cancer Institute* **91**, 226–236.

Rapp, L. & Chen, J. J. (1998). The papillomavirus E6 proteins. *Biochimica et Biophysica Acta* **1378**, F1–F19.

Rho, J., Roy-Burman, A., Kim, H., de Villiers, E. M., Matsukura, T. & Choe, J. (1994). Nucleotide sequence and phylogenetic classification of human papillomavirus type 59. *Virology* **203**, 158–161.

Tham, K. M., Chow, V. T. K., Singh, P., Tock, E. P. C., Ching, K. C., Lim-Tan, S. K., Sng, I. T. Y. & Bernard, H. U. (1991). Diagnostic sensitivity of polymerase chain reaction and Southern blot hybridization for the detection of human papillomavirus DNA in biopsy specimens from cervical lesions. *American Journal of Clinical Pathology* **95**, 638–646.

Volpers, C. & Streeck, R. E. (1991). Genome organization and nucleotide sequence of human papillomavirus type 39. *Virology* **181**, 419–423.

zur Hausen, H. & de Villiers, E. M. (1994). Human papillomaviruses. *Annual Review of Microbiology* **48**, 427–447.

Received 2 June 1999; Accepted 3 August 1999