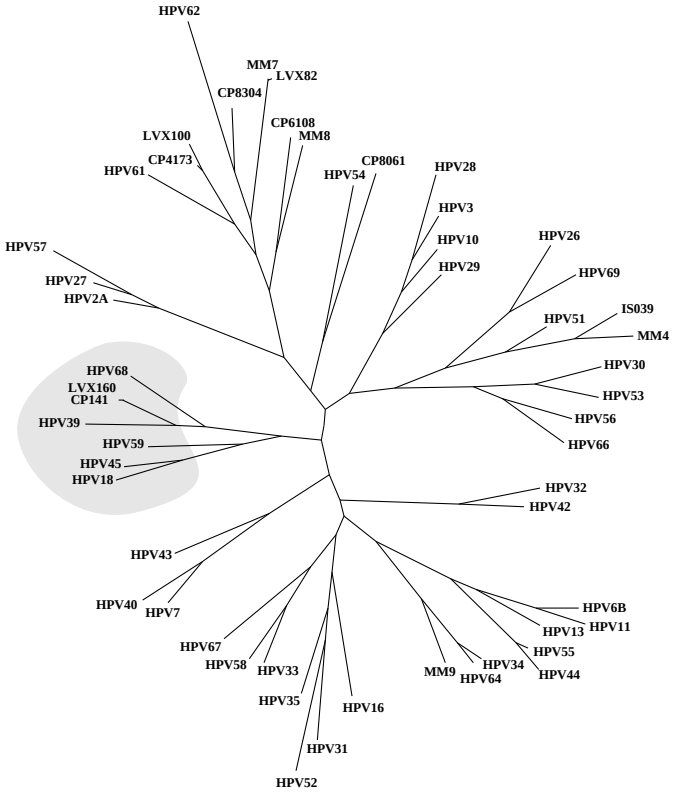
Group C Sequences

HPV18	**HPV39**
HPV45	**HPV59**
HPV68ME180	**HPVCP141**
HPVAE1	**HPVLVX160**



INTRODUCTION

Group C consists of human papillomavirus types 18, 39, 45, 59, 68ME180, and the novel viruses CP141, AE1, and LVX160, a group primarily associated with anogenital lesions, some with considerable oncogenic potential. Lorincz et al. classified HPV-18 and HPV-45 as "high risk" viruses [1]. However, Bergeron et al. placed HPV-45 in an "intermediate risk" category [2]. HPV-16, a "high risk" group A virus, was once thought to be the most lethal of the HPV types. Comparison studies of HPV-18 and HPV-16 suggest that HPV-18 may have the highest oncogenic potential of the two or, alternatively, that HPV-18 infection may progress to malignancy more rapidly. These proposals are based on research which measured the relative detection frequency of HPV-18 and HPV-16 in squamous cell-carcinomas compared to detection in CIN lesions (the ratio for HPV-16 was 1.2 as opposed to 2.3 for HPV-18) [3].

The viruses in group C primarily cause anogenital lesions. HPV-18 has been found in high prevalence in adenocarcinomas and in moderate prevalence in squamous cell carcinomas. Consistently, HPV-18 and HPV-16 are the most prevalent HPV types in adenocarcinomas and adenosquamous carcinomas [4]. Relative frequencies of the two types vary among studies. However, in most studies HPV-18 is detected at least as frequently as HPV-16 [4–7]. In contrast, HPV-18 is found less frequently than HPV-16 in squamous cell carcinomas of the genital tract, in some cases up to five times less [8, 9]. Other viruses in this group, HPV-39, HPV-68ME180, HPV-45, and HPV-59, have been isolated the majority of the time from anogenital tissues which exhibit some degree of dysplasia, if not full-blown in-situ carcinoma. HPV-39 was first isolated from penile Bowenoid papules and subsequently detected in a few cases of intraepithelial neoplasias and invasive cervical carcinomas [10]. HPV-45 was initially derived from a recurrent cervical lesion with mild to moderate dysplasia [11]. HPV-68 was originally isolated from a genital lesion [12]. Subsequently, a partial HPV genome was recovered from the cell line ME180, derived from a cervical carcinoma [13]. This sequence was more than 90% homologous to the original HPV-68 isolate [13]. And finally, HPV-59 was originally isolated from a vulvar intraepithelial neoplasm of the genital mucosa [14].

In addition to the infection of anogenital tissue, HPV-18 is highly associated with infection of the ororespiratory system. Carcinomas of the oral cavity, tongue, esophagus, sinonasal epithelium and lung have been positive for HPV-18 DNA [15, 16, 17, 18, 19]. In approximately 10% of all HPV-positive lung carcinomas HPV-18 DNA has been detected [19]. In addition, HPV-59 has been isolated from a papilloma on the lip [12].

The novel viruses CP141, AE1 and LVX160 have all been isolated from cervical samples [20, 21, 22]. CP141 has been derived from a cytologically normal cervix. Due to their recent identification, incidence rates and risk assessments have not been published for any of these novel viruses.

Of the members of Group C, complete genomic sequences are available for HPV-18, HPV-39 and HPV-45. The sequence HPV68ME180 was taken from the proviral form of the virus, and is

missing approximately 2 kbp which covers the region from the middle of E1 to the end of E2. HPV-59 and the novel sequences CP141, LVX160, and AE1 have been sequenced only over the My09-My11 fragment of L1. The sequences of CP141, LVX160 and AE1 are virtually identical to one another.

---

[1] Lorincz,A.T., Reid,R., Jenson,A.B., Greenberg,M.D., Lancaster,WD, and Kurman,R.J. Human papillomavirus infection of the cervix: relative risk associations of 15 common anogenital types. *Obestet Gynecol* **79**:328–337

[2] Bergeron,C., Barrasso,R., Beaudenon,S., Flamant,P., Croissant,O., and Orth,G. Human papillomaviruses associated with cervical intraepithelial neoplasia. Great diversity and distinct ditribution in low- and high-grade lesions. *Am J Surg Pathol* **16**:641–649

[3] Arends,M.J., Donaldson,Y.K., Duvall,E., Wyllie,A.H., and Bird,C.C. Human papillomavirus type 18 associates with more advanced cervical neoplasia than human papillomavirus type 16. *Hum Pathol* **24**:432–7 (1993).

[4] Matsuo,N., Iwasaka,T., Hayashi,Y., Hara,K., Mvula,M., and Sugimori,H. Polymerase chain reaction analysis of human papillomavirus in adenocarcinoma and adenosquamous carcinoma of the uterine cervix. *Int J Gynaecol Obstet* **41**:251–6 (1993)

[5] Leminen,A., Paavonen,J., Vesterinen,E., Wahlstrom,T., Rantala,I., and Lehtinen,M. Human papillomavirus types 16 and 18 in adenocarcinoma of the uterine cervix. *Am J Clin Pathol* **95**:647–52 (1991)

[6] Johnson,T.L., Kim,W., Plieth,D.A., and Sarkar,F.H. Detection of HPV 16/18 DNA in cervical adenocarcinoma using polymerase chain reaction (PCR) methodology. *Mod Pathol* **5**:35–40 (1992)

[7] Lee,K.R., Howard,P., Heintz,N.H., and Collins,C.C. Low prevalence of human papillomavirus types 16 and 18 in cervical adenocarcinoma in situ, invasive adenocarcinoma, and glandular dysplasia by polymerase chain reaction. *Mod Pathol* **6**:433–7 (1993)

[8] Low,S.H., Thong,T.W., Ho,T.H., Lee,Y.S., Morita,T., Singh,M., Yap,E.H., Chan,Y.C. Prevalence of human papillomavirus types 16 and 18 in cervical carcinomas: a study by dot and Southern blot hybridization and the polymerase chain reaction. *Jpn J Cancer Res* **81**:1118–23 (1990)

[9] Arends,M.J., Donaldson,Y.K., Duvall,E., Wyllie,A.H., and Bird,C.C. HPV in full thickness cervical biopsies: high prevalence in CIN 2 and CIN 3 detected by a sensitive PCR method. *J Pathol* **165**:301–9 (1991)

[10] Beaudenon,S., Kremsdorf,D., Obalek,S., Jablonska,S., Pehau-Arnaudet,G., Croissant,O., and Orth,G. Plurality of genital human papillomaviruses: characterization of two new types with distinct biological properties. *Virology* **161**:374–84 (1987)

[11] Naghashfar,Z.S., Rosenshein,N.B., Lorincz,A.T., Buscema,J., and Shah,K.V. Characterization of human papillomavirus type 45, a new type 18-related virus of the genital tract. *J Gen Virol* **68**:3073–9 (1987)

[12] de Villiers,E.M. Human pathogenic papillomavirus types: an update. in *Human pathogenic papillomaviruses*, edited by Harald zur Hausen, Springer-Verlag, Heidelberg, pp 1–12 (1994)

[13] Reuter,S., Delius,H., Kahn,T., Hofmann,B., zur Hausen,H., and Schwarz,E. Characterization of a novel human papillomavirus DNA in the cervical carcinoma cell line ME180. *J Virol* **65**:5564–8 (1991)

[14] de Villiers, E.M. Heterogeneity of the human papillomavirus group. *J Virol 63*:4898–4903 (1989)

[15] Woods,K.V., Shillitoe,E.J., Spitz,M.R., Schantz,S.P., and Adler-Storthz,K. Analysis of human papillomavirus DNA in oral squamous cell carcinomas. *J Oral Pathol Med* **22**:101–8 (1993)

[16] Toh,Y., Kuwano,H., Tanaka,S., Baba,K., Matsuda,H., Sugimachi,K., and Mori,R. Detection of human papillomavirus DNA in esophageal carcinoma in Japan by polymerase chain reaction. *Cancer* **70**:2234–8 (1992)

[17] Kashima,H.K., Kessis,T., Hruban,R.H., Wu,T.C., Zinreich,S.J., Shah, K.V. Human papillomavirus in sinonasal papillomas and squamous cell carcinoma. *Laryngoscope* **102**:973–6 (1992)

[18] Shindoh,M., Sawada,Y., Kohgo,T., Amemiya,A., and Fujinaga,K. Detection of human papillomavirus DNA sequences in tongue squamous-cell carcinoma utilizing the polymerase chain reaction method. *Int J Cancer* **50**:167–71 (1992)

[19] Snijders, P.J.F., van den Brule, A.J.C., Meijer, C.J.L.M., and Walboomers, J.M.M. Papillomaviruses and cancer of the upper digestive and respiratory tracts. in *Human pathogenic papillomaviruses*, edited by Harald zur Hausen, Springer-Verlag, Heidelberg, pp 177–197 (1994)

[20] Peyton,C.L., and Wheeler,C.M. Identification of five novel human papillomaviruses in the New Mexico triethnic population. *J. Infect. Dis.* (1994) In press

[21] Ong,C.-K., Bernard,H.-U., and Villa,L.L. Identification of genomic sequences of three novel human papillomaviruses in cervical smears of Amazonian Indians. *J. Infect. Dis.* (1994) In press

[22] Tachezy,R., Van Ranst,M.A., Cruz,Y. and Burk,R.D. Consensus primer mediated PCR allows identification of novel human papillomavirus PCR-types in cervicovaginal lavages. Unpublished

```
LOCUS       HPV18        7857 bp ds-DNA             VRL       11-DEC-1992
DEFINITION  Human papillomavirus type 18 (HPV-18), complete genome.
ACCESSION   X05015
SOURCE      Human papillomavirus type 18 DNA recovered from a cervical
            carcinoma of a Brazilian patient.
REFERENCE   1  (bases 1 to 7857)
  AUTHORS   Cole,S.T. and Danos,O.
  TITLE     Nucleotide sequence and comparative analysis of the human
            papillomavirus type 18 genome
  JOURNAL   J. Mol. Biol. 193, 599-608 (1987)
REFERENCE   2  (bases 2855-2860; revision)
  AUTHORS   Baker,C.C.
  TITLE     The Genomes of the Papillomaviruses
  JOURNAL   (in) O'Brien,S.J. (Ed.);
            Genetic Maps; Locus Maps of Complex Genomes:  1-1,
            Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1993)
COMMENT     Data kindly reviewed (14-AUG-1987) by Danos O.
            HPV-18 is most often found in lesions of the genital mucosa with
            considerable risk for malignant progression.  Estimates indicate
            that HPV-18 has been recovered from about 10-20% of all invasive
            cervical cancers.  Studies show that the predominance of HPV-18
            in high-grade anogenital lesions and invasive cancers is also
            observed in tissues of the vulva, the penis, and the anus.
            HPV-18 has been found in high prevalence in adenocarcinomas and
            in moderate prevalence in squamous cell carcinomas.   Consistently,
            HPV-18 and HPV-16 are the most prevalent HPV types in adenocarcinomas
            and adenosquamous carcinomas.  Relative frequencies of the two
            types vary among studies.  However, in most studies HPV-18 is
            detected at least as frequently as HPV-16.  In contrast, HPV-18 is
            found much less frequently than HPV-16 in squamous cell carcinomas
            of the genital tract, in some cases up to five times less.

            The 7857 bp genome of HPV-18 was originally recovered and cloned
            from a cervical carcinoma of a Brazilian patient.  This sequence
            has been corrected as stated in [2]; at nt 2855-2860 change from
            "TTGCGT" to "TGCGTT". The E7 ORF is situated immediately in front
            of E1, a characteristic common to all genital papillomaviruses
            sequenced at the time of publication. Whereas in the other subgroups,
            E7 is located in one of the other reading frames.  HPV-18 and other
            genital papillomaviruses and fibropapillomaviruses encode a hydro-
            phobic E5 gene product.  The cutaneous papillomaviruses do not
            possess a homologous E5 ORF.

            The long control region (LCR) of HPV-18 can be analyzed in three
            sections.  Segment 1 is a purine + thymidine rich area, which
            contains the polyadenylation signal for the late genes.  Segment 2
            is about 200 bp long and only appears in genital papillomaviruses.
            The third segment is the best conserved among all HPVs.  It
            contains three PV-specific palindromes, and TATA and CAAT boxes;
            the genital HPVs have one TATA box.

            The E6 and E7 ORFs contain regularly spaced cysteine doublet motifs
            with the form (Cys-X-X-Cys).  Also found in E6 of HPV-18, the
            sequence (XXXLXXXE) is found immediately after the first and
            third doublet.  Cole et al. believe these regions were derived
            from a duplication of a 33 amino acid peptide including the
            cysteine doublet.  E6 has four of these units, while E7 has three
            units, the first unit is degenerate.
BASE COUNT     2365 a   1497 c   1680 g   2315 t
ORIGIN
        1 attaatactt ttaacaattg tagtatataa aaaagggagt aACCGAAAAC GGTcgggACC
                                            E2 bind ->       E2 bind ->
       61 GAAAACGGTg tatataaaag atgTGAgaaa cacaccacaa tactATGgcg cgctttgagg
                         E6 orf start ->           E6 cds ->
                                                          |-> mRNA start site from
                                                              P(105) promoter
```

```
  121 atccaacacg gcgaccctac aagctacctg atctgtgcac ggaactgaac acttcactgc
  181 aagacataga aataacctgt gtatattgca agacagtatt ggaacttaca gaggtatttg
  241 aatttgcatt taaagattta tttgtggtgt atagagacag tataccccat gctgcatgcc
  301 ataaatgtat agatttttat tctagaatta gagaattaag acattattca gactctgtgt
  361 atggagacac attggaaaaa ctaactaaca ctgggttata caatttatta ataaggtgcc
  421 tgcggtgcca gaaaccgttg aatccagcag aaaaacttag acaccttaat gaaaaacgac
  481 gatttcacaa catagctggg cactaTAGag gccagtgcca ttcgtgctgc aaccgagcac
                     E7 orf start ->
  541 gacaggaacg actccaacga cgcagagaaa cacaagtaTA AtattaagtA TGcatggacc
                                                <- E6 end -> E7 cds
  601 taaggcaaca ttgcaagaca ttgtattgca tttagagccc caaaatgaaa ttccggttga
  661 ccttctatgt cacgagcaat taagcgactc agaggaagaa aacgatgaaa tagatggagt
  721 taatcatcaa catttaccag cccgacgagc cgaaccacaa cgtcacacaa tgttgtgtat
  781 gtgttgtaag tgtgaagcca gaattgagct agtagtagaa agctcagcag acgaccttcg
  841 agcattccag cagctgtttc tgaacaccct gtcctttgtg tgtccgtggt gtgcatccca
  901 gcagTAAgca acaATGgctg atccagaagg tacagacggg gagggcacgg gttgtaacgg
          <- E7 end
E1 orf start ->     -> E1 cds
  961 ctggttttat gtacaagcta ttgtagacaa aaaaacagga gatgtaatat cagatgacga
 1021 ggacgaaaat gcaacagaca cagggtcgga tatggtagat tttattgata cacaaggaac
 1081 attttgtgaa caggcagagc tagagacagc acaggcattg ttccatgcgc aggaggtcca
 1141 caatgatgca caagtgttgc atgttttaaa acgaaagttt gcaggaggca gcacagaaaa
 1201 cagtccatta ggggagcggc tggaggtgga tacagagtta agtccacggt tacaagaaat
 1261 atctttaaat agtgggcaga aaaaggcaaa aaggcggctg tttacaatat cagatagtgg
 1321 ctatggctgt tctgaagtgg aagcaacaca gattcaggta actacaaatg gcgaacatgg
 1381 cggcaatgta tgtagtggcg gcagtacgga ggctatagac aacgggggca cagagggcaa
 1441 caacagcagt gtagacggta caagtgacaa tagcaatata gaaaatgtaa atccacaatg
 1501 taccatagca caattaaaag acttgttaaa agtaaacaat aaacaaggag ctatgttagc
 1561 agtatttaaa gacacatatg ggctatcatt tacagattta gttagaaatt ttaaaagtga
 1621 taaaaccacg tgtacagatt gggttacagc tatatttgga gtaaacccaa caatagcaga
 1681 aggatttaaa acactaatac agccatttat attatatgcc catattcaat gtctagactg
 1741 taaatgggga gtattaatat tagccctgtt gcgttacaaa tgtggtaaga gtagactaac
 1801 agttgctaaa ggtttaagta cgttgttaca cgtacctgaa acttgtatgt taattcaacc
 1861 accaaaattg cgaagtagtg ttgcagcact atattggtat agaacaggaa tatcaaatat
 1921 tagtgaagta atgggagaca cacctgagtg gatacaaaga cttactatta tacaacatgg
 1981 aatagatgat agcaattttg atttgtcaga aatggtacaa tgggcatttg ataatgagct
 2041 gacagatgaa agcgatatgg catttgaata tgccttatta gcagacagca acagcaatgc
 2101 agctgccttt ttaaaaagca attgccaagc taaatattta aaagattgtg ccacaatgtg
 2161 caaacattat aggcgagccc aaaaacgaca aatgaatatg tcacagtgga tacgatttag
 2221 atgttcaaaa atagatgaag ggggagattg gagaccaata gtgcaattcc tgcgatacca
 2281 acaaatagag tttataacat ttttaggagc cttaaaatca tttttaaaag gaacccccaa
 2341 aaaaaattgt ttagtatttt gtggaccagc aaatacagga aaatcatatt ttggaatgag
 2401 ttttatacac tttatacaag gagcagtaat atcatttgtg aattccacta gtcatttttg
 2461 gttggaaccg ttaacagata ctaaggtggc catgttagat gatgcaacga ccacgtgttg
 2521 gacatacttt gatacctata tgagaaatgc gttagatggc aatccaataa gtattgatag
 2581 aaagcacaaa ccattaatac aactaaaatg tcctccaata ctactaacca caaatataca
                      |-> mRNA start site from P(2598) promoter
 2641 tccagcaaag gataatagat ggccatattt agaaagtaga ataacagtat ttgaatttcc
 2701 aaatgcattt ccatttgata aaaatggcaa tccagtatat gaaataaatg acaaaaattg
 2761 gaaatgtttt tttgaaagga catggtccag atTAGatttg cacgaggaag aggaagATGc
                             E2 orf start ->             E2 cds ->
 2821 agacaccgaa ggaaaccctt tcggaacgtt taagtgcgtt gcaggacaaa atcatagacc
 2881 actaTGAaaa tgacagtaaa gacatagaca gccaaataca gtattggcaa ctaatacgtt
          <- E1 end
 2941 gggaaaatgc aatattcttt gcagcaaggg aacatggcat acagacatta aaccaccagg
 3001 tggtgccagc ctataacatt tcaaaaagta aagcacataa agctattgaa ctgcaaatgg
                                          |-> mRNA start site from
                                              P(3036) promoter
 3061 ccctacaagg ccttgcacaa agtcgataca aaaccgagga ttggacactg caagacacat
 3121 gcgaggaact atggaataca gaacctactc actgctttaa aaaaggtggc caaacagtac
 3181 aagtatattt tgatggcaac aaagacaatt gtatgaccta tgtagcatgg gacagtgtgt
 3241 attatatgac tgatgcagga acatgggaca aaaccgctac ctgtgtaagt cacaggggat
 3301 tgtattatgt aaaggaaggg tacaacacgt tttatataga atttaaaagt gaatgtgaaa
 3361 aatatgggaa cacaggtacg tgggaagtac attttgggaa taatgTAAtt gattgtaATG
```

```
                                              E4 orf start -> E4 cds ->
3421 actctatgtg cagtaccagt gacgacacgg tatccgctac tcagcttgtt aaacagctac
3481 agcacacccc ctcaccgtat tccagcaccg tgtccgtggg caccgcaaag acctacggcc
3541 agacgtcggc tgctacacga cctggacact gtggactcgc ggagaagcag cattgtggac
3601 ctgtcaaccc acttctcggt gcagctacac ctacaggcaa caacaaaaga cggaaactct
3661 gtagtggtaa cactacgcct aTAAtacatt taaaaggtga cagaaacagt ttaaaatgtt
                            <- E4 end
3721 tacggtacag attgcgaaaa catagcgacc actatagaga tatatcatcc acctggcatt
3781 ggacaggtgc aggcaatgaa aaaacaggaa tactgactgt aacataccat agtgaaacac
3841 aaagaacaaa atttttaaat actgttgcaa ttccagatag tgtacaaata ttggtgggat
3901 acatgacaat gTAAtacata tgctgtagta ccaatATGtt atcacttatt tttttatttt
                  <- E2 end         E5 cds ->
       E5 orf start ->
3961 gcttttgtgt atgcatgtat gtgtgctgcc atgtcccgct tttgccatct gtctgtatgt
4021 gtgcgtatgc atgggtattg gtatttgtgt atattgtggt aataacgtcc cctgccacag
4081 cattcacagt atatgtattt tgtttttat tgcccatgtt actattgcat atacatgcta
4141 tattgtcttt acagTAAttg taTAGgttgt tttatacagt gtattgtaca ttgtatattt
                    <- E5 end
                   L2 orf start ->
4201 tgttttatac cttttatgct ttttgtattt ttgtaataaa agtATGgtat cccaccgtgc
                                             L2 cds ->
4261 cgcacgacgc aaacgggctt cggtaactga cttatataaa acatgtaaac aatctggtac
4321 atgtccacct gatgttgttc ctaaggtgga gggcaccacg ttagcagata aaatattgca
4381 atggtcaagc cttggtatat ttttgggtgg acttggcata ggtactggca gtggtacagg
4441 gggtcgtaca gggtacattc cattgggtgg gcgttccaat acagtggtgg atgttggtcc
4501 tacacgtccc ccagtggtta ttgaacctgt gggccccaca gacccatcta ttgttacatt
4561 aatagaggac tccagtgtgg ttacatcagg tgcacctagg cctacgttta ctggcacgtc
4621 tgggtttgat ataacatctg cgggtacaac tacacctgcg gttttggata tcacaccttc
4681 gtctacctct gtgtctattt ccacaaccaa ttttaccaat cctgcatttt ctgatccgtc
4741 cattattgaa gttccacaaa ctggggaggt ggcaggtaat gtatttgttg gtacccctac
4801 atctggaaca catgggtatg aggaaatacc tttacaaaca tttgcttctt ctggtacggg
4861 ggaggaaccc attagtagta ccccattgcc tactgtgcgg cgtgtagcag gtccccgcct
4921 ttacagtagg gcctaccaac aagtgtcagt ggctaaccct gagtttctta cacgtccatc
4981 ctctttaatt acatatgaca acccggcctt tgagcctgtg gacactacat taacatttga
5041 tcctcgtagt gatgttcctg attcagattt tatggatatt atccgtctac ataggcctgc
5101 tttaacatcc aggcgtggga ctgttcgctt tagtagatta ggtcaacggg caactatgtt
5161 tacccgcagc ggtacacaaa taggtgctag ggttcacttt tatcatgata taagtcctat
5221 tgcaccttcc ccagaatata ttgaactgca gcctttagta tctgccacgg aggacaatga
5281 cttgtttgat atatatgcag atgacatgga ccctgcagtg cctgtaccat cgcgttctac
5341 tacctccttt gcatttttta aatattcgcc cactatatct tctgcctctt cctatagtaa
5401 tgtaacggtc ccttTAAcct cctcttgggA TGtgcctgta tacacgggtc ctgatattac
         L1 orf start ->    L1 cds ->
5461 attaccatct actacctctg tatggcccat tgtatcaccc acggcccctg cctctacaca
5521 gtatattggt atacatggta cacattatta tttgtggcca ttatattatt ttattcctaa
5581 gaaacgtaaa cgtgttccct atttttttgc agatggcttt gtggcggccT AGtgacaata
                        |-> mRNA start site from P(5600) promoter
                                                              <- L2 end
5641 ccgtatatct tccacctcct tctgtggcaa gagttgtaaa taccgatgat tatgtgactc
5701 ccacaagcat attttatcat gctggcagct ctagattatt aactgttggt aatccatatt
5761 ttagggttcc tgcaggtggt ggcaataagc aggatattcc taaggtttct gcataccaat
5821 atagagtatt tagggtgcag ttacctgacc caaataaatt tggtttacct gatactagta
5881 tttataatcc tgaaacacaa cgtttagtgt gggcctgtgc tggagtggaa attggccgtg
5941 gtcagccttt aggtgttggc cttagtgggc atccatttta taataaatta gatgacactg
6001 aaagttccca tgccgccacg tctaatgttt ctgaggacgt tagggacaat gtgtctgtag
6061 attataagca gacacagtta tgtattttgg gctgtgcccc tgctattggg gaacactggg
6121 ctaaaggcac tgcttgtaaa tcgcgtcctt tatcacaggg cgattgcccc cctttagaac
6181 ttaaaaacac agttttggaa gatggtgata tggtagatac tggatatggt gccatggact
6241 ttagtacatt gcaagatact aaatgtgagg taccattgga tatttgtcag tctatttgta
6301 aatatcctga ttatttacaa atgtctgcag atccttatgg ggattccatg ttttttgct
6361 tacggcgtga gcagcttttt gctaggcatt tttggaatag agcaggtact atgggtgaca
6421 ctgtgcctca atccttatat attaaaggca caggtatgcc tgcttcacct ggcagctgtg
6481 tgtattctcc ctctccaagt ggctctattg ttacctctga ctcccagttg tttaataaac
6541 catattggtt acataaggca cagggtcata acaatggtgt ttgctggcat aatcaattat
6601 ttgttactgt ggtagatacc actcccagta ccaatttaac aatatgtgct tctacacagt
```

```
6661 ctcctgtacc tgggcaatat gatgctacca aatttaagca gtatagcaga catgttgagg
6721 aatatgattt gcagtttatt tttcagttgt gtactattac tttaactgca gatgttatgt
6781 cctatattca tagtatgaat agcagtattt tagaggattg gaactttggt gttccccccc
6841 ccccaactac tagtttggtg gatacatatc gttttgtaca atctgttgct attacctgtc
6901 aaaaggatgc tgcaccggct gaaaataagg atccctatga taagttaaag ttttggaatg
6961 tggatttaaa ggaaaagttt tctttagact tagatcaata tccccttgga cgtaaatttt
7021 tggttcaggc tggattgcgt cgcaagccca ccataggccc tcgcaaacgt tctgctccat
7081 ctgccactac gtcttctaaa cctgccaagc gtgtgcgtgt acgtgccagg aagTAAtatg
                                                          <- L1 end
7141 tgtgtgtgta tatatatata catctattgt tgtgtttgta tgtcctgtgt ttgtgtttgt
7201 tgtatgattg cattgtatgg tatgtatggt tgttgttgta tgttgtatgt tactatattt
7261 gttggtatgt ggcattaaat aaaatatgtt ttgtggttct gtgtgttatg tggttgcgcc
7321 ctagtgagta acaactgtat ttgtgtttgt ggtatgggtg ttgcttgttg ggctatatat
7381 tgtcctgtat ttcaagttat aaaactgcac accttacagc atccatttta tcctacaatc
7441 ctccattttg ctgtgcaACC GATTTCGGTt gcctttggct tatgtctgtg gttttctgca
              E2 bind ->
7501 caatacagta cgctggcact attgcaaact ttaatctttt gggcactgct cctacatatt
7561 ttgaacaatt ggcgcgcctc tttggcgcat ataaggcgca cctggtatta gtcattttcc
7621 tgtccaggtg cgctacaaca attgcttgca taactatatc cactccctaa gtaataaaac
7681 tgcttttagg cacatatttt agtttgtttt tacttaagct aattgcatac ttggcttgta
7741 caactacttt catgtccaac attctgtcta cccttaacat gaactataat atgactaagc
7801 tgtgcataca tagtttatgc aACCGAAATA GGTtgggcag cacatactat acttttc
                   E2 bind ->
```

```
LOCUS       HPV39        7833 bp ds-DNA             VRL       06-MAR-1991
DEFINITION  Human papillomavirus type 39 (HPV-39), complete genome.
ACCESSION   M62849 M38185
SOURCE      Human papillomavirus type 39 DNA isolated from a penile Bowenoid
            papule biopsy.
REFERENCE   1  (bases 1 to 7833)
  AUTHORS   Volpers,C. and Streeck,R.E.
  TITLE     Genome organization and nucleotide sequence of human papillomavirus
            type 39
  JOURNAL   Virology 181, 419-423 (1991)
COMMENT     HPV-39 is most often found in lesions of the genital mucosa which
            have a risk for malignant progression.  Beaudenon et al. conducted
            a study which sceened 365 HPV positive patients for HPV-39 DNA,
            the resultant detection rate was 3.9%.

            The 7833 bp genome of HPV-39 was first recovered and cloned from
            penile Bowenoid papules.  It was subsequently detected in a few cases
            of intraepithelial neoplasias and invasive cervical carcinomas. The
            physical state of the DNA was determined to be episomal.  The genome
            contains an E7 ORF which is located immediately upstream of E1, common
            among all genital papillomaviruses.  Unusually, a large ORF of 1.3 kb
            has been found on the complementary strand.  This ORF contains an
            initiation codon, a potential splice acceptor site close to the
            5' end, and a polyadenylation signal at the 3' end.  Further upstream
            of this large ORF is a smaller ORF preceeded by a TATA box and a NF-1
            binding site.

            The noncoding region of HPV-39 contains several features common
            to other papillomavirus types.  It contains three complete and
            two degenerate versions of an E2 binding site.  Possible promoter
            elements which have been identified include two TATA boxes, a
            conserved AAAGGGAGTA promoter element which is upstream of a 12 bp
            palindrome tandem repeat, and an enhancer core sequence.  Various
            transcription factor binding sites are also present.  These include
            four possible sites for nuclear factor 1 (NF-1), two possible sites
            for activator protein 1 (AP-1), and a motif for the papillomavirus
            enhancer associated factor (PVF).  A glucocorticoid response element
            (GRE) is found resembling those found in other types.  In addition,
            a GRE is found in the L1 ORF with no equivalent in other types.

            The E6 and E7 ORFs of HPV-39 contain four copies and one copy
            respectively of the well-conserved cysteine doublet (Cys-X-X-Cys)
            motif.  Mutational analysis of the HPV-16 ORF has shown that one
            copy of this motif is sufficient for transformation.  In addition,
            the E7 ORF of HPV-39 contains a putative cell division motif found
            in genital HPVs associated with malignancy, SV40 large T antigen,
            adenovirus E1A, and the myc protein.

BASE COUNT     2426 a   1485 c   1660 g   2262 t
ORIGIN
        1 cttataacat tttataagta tcttgtttaa aaAAAGGGAG TAACCGAAAA CGGTcaggAC
                                  promoter ->       <--> E2 bind       ->
                                           E6 orf start ->
       61 CGAAATCGGT ggaTATAAAA cgcagtcaca gtttctgtcc ataccgATGg cgcgatttca
          E2 bind        -> signal                    E6 cds ->
      121 caatcctgca gaacggccat acaaattgcc agacctgtgc acaacgctgg acaccacctt
      181 gcaggacatt acaatagcct gtgtctattg cagacgacca ctacagcaaa ccgaggtata
      241 tgaatttgca tttagtgatt tatatgtagt atatagggac ggggaaccac tagctgcatg
      301 ccaatcatgt ataaaatttt atgctaaaat acgggagcta cgatattact cggactcggt
      361 gtatgcaact acattagaaa atataactaa tacaaagtta tataatttat taataaggtg
      421 catgtgttgt ctgaaaccgc tgtgtccagc agaaaaatta agacacctaa atagcaaacg
      481 aagatttcaT AAaatagcag gaagctatac aggacagtgt cgacggtgct ggaccacaaa
          E7 orf start ->
      541 acgggaggac cgcagactaa cacgaagaga aacccaagta TAAcatcaga tATGcgtgga
                                                    <- E6 end  -> E7 cds
      601 ccaaagccca ccttgcagga aattgtatta gatttatgtc cttacaatga aatacagccg
```

```
 661 gttgaccttg tatgtcacga gcaattagga gagtcagagg atgaaataga tgaacccgac
 721 catgcagtta atcaccaaca tcaactacta gccagacggg atgaaccaca gcgtcacaca
 781 atacagtgtt cgtgttgtaa gtgtaacaac acactgcagc tggtagtaga agcctcacgg
 841 gatactctgc gacaactaca gcagctgttt atggactcac taggatttgt gtgtccgtgg
 901 tgtgcaactg caaaccagTA AcctgctATG gccaatcgtg aaggtacaga cggggatggg
               E1 orf start ->    -> E1 cds
                          <- E7 end
 961 tcgggatgta acggatggtt tctagtacag gcaatagtag ataaacaaac aggcgacaca
1021 gtgtcggagg atgaggatga aaatgcaaca gatacaggtt cagacctggc agactttatt
1081 gatgattcca cagatatttg tgtacaggca gagcgtgaga cagcacaggt acttttacat
1141 atgcaagagg cccaaaggga tgcacaagca gtgcgtgcct taaaacgaaa gtatacagac
1201 agcagtggcg acactagacc gtatggaaaa aaagtaggca ggaataccag gggaacacta
1261 caggaaattt cattaaatgt aagcagtacg caggcaacac aaacggtgta ttccgtgcca
1321 gacagcggat atggcaatat ggaagtggaa acagctgaag tggaggaggt aactgtagca
1381 actaatacaa atggggatgc tgaaggggaa catggcggca gtgtacggga ggagtgcagt
1441 agtgtggata gtgctataga tagtgaaaac caggatccca aatctccaac tgcacaaatt
1501 aaattattgt tacaatccaa taacaaaaag gctgcaatgc taacacaatt taaagaaaca
1561 tatggactat cctttactga cctggtacgt acgtttaaaa gtgataaaac aacatgtaca
1621 gactgggtgg cagccatatt tggagtacat ccaactattg cagaaggatt taaaacatta
1681 atcaacaaat atgccttata tacacatata caaagcttag acacaaaaca aggagtacta
1741 attttaatgc taataagata tacatgtgga aaaaataggg ttactgtagg aaagggatta
1801 agtacattgt tacatgttcc agaaagttgt atgcttctgg agcctcctaa actgcgcagc
1861 cctgtagcag cactatattg gtatcgcaca ggtatatcca atattagtgt ggtaacaggg
1921 gatacgccag aatggataca acgattaact gttatacaac atggaataga tgatagtgta
1981 tttgacctat cggacatggt acaatgggca tttgacaatg aatatactga tgaaagtgac
2041 atagcattta attatgcaat gttagcagat tgtaacagta atgctgcagc ctttttaaaa
2101 agtaactgcc aggcaaaata tgtaaaagat tgtgcaacaa tgtgtaaaca ttacaagcga
2161 gcacaaaaaa ggcaaatgtc catgtctcaa tggataaaat ttaggtgtag taaatgtgat
2221 gaaggcgggg actggagacc catagtacaa ttcttaagat atcaaggaat agaatttata
2281 tccttttttat gtgcattaaa ggaatttta aagggtactc ccaaaaaaaa ctgtatagtt
2341 atatatggac ctgcgaatac aggaaagtca cattttgta tgagccttat gcatttttta
2401 cagggcacag ttatttcata tgtaaactcc accagccact tttggctaga accacttgca
2461 gatgcaaaac tagcaatgtt agatgatgca accggtacct gctggtcata tttcgataat
2521 tatatgagaa atgcattaga tgggtatgca ataagtttag ataggaaata taaaagttta
2581 ctacaaatga aatgtccacc attattaata acctccaata ccaatcctgt ggaagacgat
2641 aggtggccat atttacgtag taggctaaca gtgtttaaat ttcctaatgc atttccattt
2701 gaccaaaaca ggaatccagt gtacacaatc aatgataaaa actggaaatg ttttttgaa
2761 aagacttggt gcagatTAGa cttgcagcag gacgaggATG aaggagacaa tgatgaaaac
           E2 orf start ->           E2 cds ->
2821 actttcacaa cgtttaaatg tgttacagga caaaatacta gaatactaTG Aacaagacag
                                                          <- E1 end
2881 taaatcaata tatgatcaaa ttaattattg gaaatgtgtg cgaatggaaa atgcaatatt
2941 ttatgcagca cgagaacgtg gcatgcatac tattgaccac caggtggtgc caaccataaa
3001 catttcaaaa tgtaaagcat atcaagctat tgaactgcag atggcactag aaagtgttgc
3061 acaaactgaa tacaatacag aggagtggac attaaaagac actagtaatg aactgtggca
3121 tacacagcca aaacaatgtt ttaaaaaaca aggaactaca gtggaggtgt ggtatgatgg
3181 ggacaaatgt aatgctatga actatgtatt atggggtgct atatattata aaaataatat
3241 agacatatgg tgtaaaacag aagggtgtgt ggactattgg ggtatatatt atatgaacga
3301 gcacctaaaa gtatactatg aagtgtttat tcaagatgcg gaaaggtatg ggactagtgg
3361 caaatgggaa gtgcattata atggcaacaT AAttcattgt cctgactcta tgtgcagtac
                          E4 orf start ->
                          NH2 terminus unknown
3421 cagtgacgga tcggtaccca ctactgaact tactaccgaa ttatcaaaca ccaccgcgac
3481 ccattccacc gcaacaaccc catgcaccca aaaaacaatc ccgccgccgt ctcgaaagcg
3541 acctcgacag tgtgcagtca cagagcccac tgagcccgac ggagtgtccc tggaccatct
3601 taacaaccca ctccacagta acagtacagg ccacaacaca agacggtacc tcagttgtgg
3661 taacactacg cctaTAAtac atttaaaagg tgacaaaaat ggtttaaaat gtttaagata
                   <- E4 end
3721 tagactacaa aaatatgaca cattgtttga aaatatttca tgtacctggc attggatacg
3781 gggtaaggga accaaaaacg ctggcatatt aactgttaca tatgccacag agtcacaacg
3841 ccaaaaattt ttggacactg ttaaaatacc ttctagtgta catgtttcat tgggttacat
3901 gacattgTAA agtatactat ggatattgtg tatgtatatt gtatacatac tacaTAGATG
            <- E2 end                                     E5 cds ->
3961 atattattgg tatttttggt gtggtttggt gtgtgtatat atatatgttg caatgtcccg
```

```
4021 cttttgccgt ctgtgcatgt gtgtgcgtat gtgtggataa ttgtgtttgt gtttattctt
4081 atacgtacca caccattgga ggtgtttttt gtatatttac tatttttgt attgcccatg
4141 tggttgttgc atagactggc aatggataTG AtaTAGtact gtatatgtat gtgcattgtg
                         L2 orf start -> <- E5 end
4201 cataactact gtacatagct ttttatattt ttttttgtta ctAATAAAcA TGgtttccca
                                            signal ->
                                                   L2 cds ->
4261 ccgtgctgcc aggcgtaagc gtgcatctgc aactgaccta tatagaacct gtaaacaatc
4321 gggtacctgt ccaccagacg ttgttgataa agttgagggt actacacttg ctgacaaaat
4381 tttacagtgg actagtttag gtatattttt gggtgggtta ggcataggca caggtactgg
4441 tactgggggga cgcacaggat atatacccct ggggggtagg cctaatactg ttgtagatgt
4501 gtctcctgca cgtccacctg tagttattga acctgttggt ccttctgagc catctattgt
4561 gcaattggtg gaggactcaa gtgttataac ctctggaaca ccagtaccaa catttacagg
4621 cacctctgga tttgaaatta cttcttcttc tactactacg cctgcggtat tggatattac
4681 accctcctct gggtctgtac aaataacctc tactagttat actaaccctg cctttacgga
4741 tccttcctta attgaggttc cccaaacagg tgaaacctcg ggtaatatat ttgtcagtac
4801 ccctacatca ggtacacatg gctatgagga aatacctatg gaagtgtttg ccacacatgg
4861 cacaggtacc gaacctatta gcagcacacc tacacctgga atcagtcgtg tggcaggacc
4921 acgtttatat agtagagcac atcagcaggt tcgtgttagt aattttgatt ttgtaactca
4981 cccttcatca tttgtaacat ttgataatcc tgcttttgag cctgttgata ctacattaac
5041 atatgaagct gctgacatag ctccagatcc ggattttctg gacattgttc gtttacatag
5101 gcctgcctta acctcgcgta aaggaacagt aaggtttagt aggcttggca aaaaggctac
5161 catggttacc cggcgtggca cacaaattgg agcgcaagta cattattacc atgacattag
5221 tagtattgct cctgctgaaa gcattgaatt acagccccta gttcacgctg agccctctga
5281 tgcttcagat gcattatttg atatatatgc tgatgtggac aataacacat atttagatac
5341 tgcatttaat aatacaaggg attcgggcac tacatataac acaggctcac taccttctgt
5401 ggcttcttca gcatctacta aatatgccaa tacaactatt ccttttagta cctcatggaa
5461 tatgcctgta aatactggtc ctgatattgc tttaccaagt actactccac agttgccatt
5521 ggtgccttct ggaccaatag acacaacata tgcaataacc attcagggtt ccaattatta
5581 tttgttgcca ttattgtatt ttttccTAAa aaaacgtaaa cgtattccct attttttttc
                       L1 orf start ->
5641 agATGgctat gtggcggtcT AGtgacagca tggtgtattt gcctccacct tctgtggcga
L1 cds ->                  <- L2 end
5701 aggttgtcaa tactgatgat tatgttacac gcacaggcat atattattat gctggcagct
5761 ctagattatt aacagtagga catccatatt ttaaagtggg tatgaatggt ggtcgcaagc
5821 aggacattcc aaaggtgtct gcatatcaat atagggtatt tcgcgtgaca ttgcccgatc
5881 ctaataaatt cagtattcca gatgcatcct tatataatcc agaaacacaa cgtttagtat
5941 gggcttgtgt aggggtggag gtgggcaggg gccagccatt gggtgttggt attagtggac
6001 acccattata taatagacag gatgatactg aaaactcacc attttcatca accaccaata
6061 aggacagtag ggataatgtg tctgtggatt ataaacagac acagttgtgc attataggct
6121 gtgttcccgc cattggggag cactggggta agggaaaggc atgcaagccc aataatgtat
6181 ctacggggga ctgtcctcct ttggaactag taaacacccc tattgaggat ggtgatatga
6241 ttgatactgg ctatggagct atggactttg gtgcattgca ggaaaccaaa agtgaggtgc
6301 ctttagatat ttgtcaatcc atttgtaaat atcctgatta tttgcaaatg tctgcagatg
6361 tgtatgGGGA CAGTATGTTC Ttctgtttac gtagggaaca actgtttgca agacattttt
           -> glucocorticoid response element
6421 ggaatcgtgg tggtatggtg ggtgacgcca ttcctgccca attgtatatt aagggcacag
6481 atatacgtgc aaaccccggt agttctgtat actgcccctc tcccagcggt tccatggtaa
6541 cctctgattc ccagttattt aataagcctt attggctaca taaggcccag ggccacaaca
6601 atggtatatg ttggcataat caattatttc ttactgttgt ggacactacc cgtagtacca
6661 actttacatt atctacctct atagagtctt ccataccttc tacatatgat ccttctaagt
6721 ttaaggaata taccaggcac gtggaggagt atgatttaca atttatattt caactgtgta
6781 ctgtcacatt aacaactgat gttatgtctt atattcacac tatgaattcc tctatattgg
6841 acaattggaa ttttgctgta gctcctccac catctgccag tttggtagac acttacagat
6901 acctacagtc tgcagccatt acatgtcaaa aggatgctcc agcacctgaa aagaaagatc
6961 catatgacgg tctaaagttt tggaatgttg acttaaggga aaagtttagt ttggaacttg
7021 atcaattccc tttgggacgt aaatttttgt tgcaggccag ggtccgcagg cgccctacta
7081 taggtccccg aaagcggcct gctgcatcca cttcctcgtc ctcagctact aaacacaaac
7141 gtaaacgtgt gtctaaaTAA tgcatgtgta tgccttgtta tgtgtgtgta tgttgtttgt
                        <- L1 end
7201 ttccttatgt gttgagtgta tatgtgtatg tttgtaggta tgtgtgtata tgtttttgtt
7261 AATAAAgtat gtatgacagt ttcatgtgtg attgcacacc ctgtgactaa cagtgtattt
     -> signal
7321 gttttacata taataggtct gcaacatttc atacataatc tatatgccct accctaaggt
```

```
7381 gtgtttacta cctaatatgt aatttttaca ttgttgtatg cgttTCTACA TTTTATACTt
                     glucocorticoid response element ->
7441 cgccattttg tggcgACCGA AGTCGGTcgt gggttgagca ttttttttaa actagtggaa
             E2 bind ->
7501 accacctttc tcagcaaaaa catgtcttta ccttaggttc accctgcata gttggcactg
7561 gtaacagttt tactggcgcg ccttattact catcatcctg tccaggtgca ctgcaacaat
7621 actttggcaa catccatatc tccaccctat gtaataaaac tgcttttagg catatatttt
7681 agctgttttt acttgcttaa ttaaatagtt ggcctgtata actacttttt gattcaggaa
7741 tgtgtcttac agtataagtt atacaagtga ctaatgtagc acacaatagt ttatgcaACC
                                                                    ->
7801 GAAATAGGTt gggcatacat acctatactt tta
    E2 bind
```

HPV45

```
LOCUS       HPV45        7858 bp ds-DNA             VRL       04-OCT-1993
DEFINITION  Human papillomavirus type 45 (HPV-45), complete genome.
ACCESSION   X74479
SOURCE      Human papillomavirus type 45 DNA.
REFERENCE   1  (bases 1 to 7858)
  AUTHORS   Delius,H. and Hofmann,B.
  TITLE     Primer-directed sequencing of human papillomavirus types
  JOURNAL   Curr. Top. Microbiol. Immunol. 186, 13-31 (1994)
REFERENCE   2  (bases 1 to 7858)
  AUTHORS   Delius,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-1993) to the EMBL/GenBank/DDBJ databases. H.
            Delius, Deutsches Krebsforschungszentrum, Abteilung ATV, Im
            Neuenheimer Feld 506, W 6900 Heidelberg, FRG
COMMENT     Lorincz et al. (Obestet Gynecol 79:328-337) classified HPV-45 as
            a "high risk" virus.  DNA from HPV-45, as well as from others
            in the high-risk class, was detected in 7% of the low-grade cervical
            lesions, 7% of the high-grade lesions, and in 27% of invasive
            cancers screened.  However, conflicting data placed HPV-45 in an
            "intermediate risk" category (Bergeron, C., et al. Am J Surg Pathol
            16:641-649). HPV-45 was initially derived from a recurrent cervical
            lesion with mild to moderate dysplasia (Naghashfar et al. J Gen
            Virol 68:3073-9) and subsequently sequenced by Dr. H. Delius.  In
            initial prevalence studies, HPV-45 was detected in only 3 cases out
            of roughly 600 genital tissues tested.
BASE COUNT     2409 a   1462 c   1652 g   2335 t
ORIGIN      101 bp upstream from beginning of E6 cds
        1 aatactttta acaattatac tacataaaaa agggtgtaAC CGAAAACGGT tgcaACCAAA
                                                 -> E2 bind        -> E2 bind
       61 AACGGTgcat aTAAaagctt tgtggaaaag tgcattacag gATGgcgcgc tttgacgatc
            E6 orf start ->                    E6 cds ->
      121 caaagcaacg accctacaag ctaccagatt tgtgcacaga attgaataca tcactacaag
      181 acgtatctat tgcctgtgta tattgcaaag caacattgga acgcacagag gtatatcaat
      241 ttgcttttaa agatttatgt atagtgtata gagactgtat agcatatgct gcatgccata
      301 aatgtataga cttttattcc agaattagag aattaagata ttattcaaac tctgtatatg
      361 gagagacact ggaaaaaata actaatacag agttgtataa tttgttaata aggtgcctgc
      421 ggtgccagaa accattgaac ccagcagaaa aacgtagaca ccttaaggac aaacgaagat
      481 ttcacagcat agctggacag taccgagggc agtgtaatac atgttgtgac caggcacggc
      541 aagaaagact tcgcagacgT AGggaaacac aagtaTAGca ataagtATGc atggaccccg
                  E7 orf start ->                E7 cds ->
                                               <- E6 end
      601 ggaaacactg caagaaattg tattgcattt ggaacctcag aatgaattag atcctgttgA
                                                                        ->
      661 CCTGTTGTGT Tacgagcaat taagcgagtc agaggaggaa aacgatgaag cagatggagt
            E2 bind
      721 tagtcatgca caactaccag cccgacgagc cgaaccacag cgtcacaaaa ttttgtgtgt
      781 atgttgtaag tgtgacggca gaattgagct tacagtagag agctcggcag aggaccttag
      841 aacactacag cagctgtttt tgagcacctt gtcctttgtg tgtccgtggt gtgcaactaa
      901 ccaaTAAtct acaATGgcgg atccagaagg taccgacggg gagggaacgg ggtgtaatgg
                  E1 cds ->
    E1 orf start ->
               <- E7 end
      961 ctggttcttt gtagaaacaa ttgtagagaa aaaaacaggg gatgtaatat cagatgatga
     1021 ggatgaaact gcaacagata cagggtcgga tatggtagat tttattgaca cacaattatc
     1081 catttgtgaa caggcagagc aagagacagc acaggcattg ttccatgcgc aggaagttca
     1141 gaatgatgca caggtgttgc atcttttaaa acgaaagttt gcaggaggca gcaaggaaaa
     1201 cagtccatta ggggagcagc taagtgtgga tacggatcta agtccacggt tacaagaaat
     1261 ttcattaaat agtgggcaca aaaaagcaaa acgacggttg tttacaatat cagatagtgg
     1321 ctatggctgt tctgaagtgg aagctgcaga gactcaggta actgtaaaca ctaatgcgga
     1381 aaatggcggc agtgtacata gtacacaaag tagtggtggg gatagtagtg acaatgcaga
     1441 aaatgtagat ccgcattgca gtattacaga actaaaggag ctattacaag caagtaacaa
     1501 aaaggctgca atgctggcag tatttaaaga catatatggg ctgtcattta cggatttggt
     1561 tagaaatttt aaaagtgata aaacaacatg tacagattgg gtaatggcta tatttggagt
     1621 taatccaacg gtagcagaag gctttaaaac attaattaaA CCAGCAACGT Tatacgccca
```

```
                                                  -> E2 bind
  1681 tatccaatgt ttagattgta aatggggagt attaatatta gctttattaa gatataaatg
  1741 tggcaaaaat agactaactg ttgcaaaagg cttaagcaca ttgttgcacg tacctgaaac
  1801 atgtatgtta attgaaccac caaaattgcg aagtagtgtt gcagcattat actggtatag
  1861 aacaggtata tccaatatta gtgaagtaag tggagacaca cctgagtgga tacaaagact
  1921 gacaattatt caacatggta ttgacgatag taattttgat ttgtcagaca tggtgcaatg
  1981 ggcatttgat aatgacctta cagatgaaag tgatatggca tttcaatatg cccaattagc
  2041 agactgcaac agtaatgcag ctgcattttt aaaaagtaac tgccaagcca aatatttaaa
  2101 agattgtgct gtaatgtgta gacattataa aagagcacaa aaacgccaaa tgaatatgtc
  2161 tcaatggatt aaatatagat gttccaaaat agatgaaggt ggggattgga gacccatagt
  2221 acaattccta agatatcagg gagtagaatt tattagcttt ttaagggcac taaaggaatt
  2281 tcttaaagga acaccaaaaa aaaattgtat actgttatat ggacctgcaa atacaggaaa
  2341 atcgtatttt ggaatgagtt ttatacattt cctacaaggt gcaataatat catttgtaaa
  2401 ttcaaacagc catttttggt tagaaccgtt agcagatact aaggtagcca tgttggatga
  2461 tgccacacac acgtgttgga catattttga taattatatg agaaatgcat tagatggtaa
  2521 tcctataagt atagacagaa agcataaacc attattacag ctaaaatgtc ctccaatcct
  2581 attaacatcc aatattgatc cagcaaaaga taataaatgg ccatatttag aaagtagggt
  2641 gacggtattt acatttccac atgcatttcc atttgataaa aatggtaatc cagtatatga
  2701 aataaatgat aaaaattgga aatgtttttt tgaaaggaca tggtccagat TAGatttgca
                                                    E2 orf start ->
  2761 cgaggacgAT Gaagatgcag acaccgaagg aatccctttc ggaacgttta agtgcgttac
       E2 cds ->
  2821 aggacaaaat actagaccac taTGAaaatg acagtaaaga cataaacagc caaataagtt
                              <- E1 end
  2881 attggcaact tatacgtttg gaaaatgcaa tactatttac agcaagggaa catggtatta
  2941 ccaaactaaa ccaccaggtg gtgcctccta ttaacatttc aaaaagcaaa gcacataaag
  3001 ctattgaact gcaaatggcc ttaaagggcc ttgcacaaag caagtataac aatgaggaat
  3061 ggacactgca agatacatgc gaggaactat ggaatacaga accgtcgcag tgttttaaaa
  3121 aaggcggtaa aaccgtgcac gtatactttg atggcaacaa ggacaactgt atgaactatg
  3181 tagtatggga cagtatatat tatataactg agacagggat atgggacaaa acagcagcat
  3241 gtgttagcta ttggggtgta tattatataa aagatggaga taccacatat tatgtacaat
  3301 ttaaaagcga atgtgagaaa tatggaaata gtaatacgtg ggaagtacaa tatgggggca
  3361 atgTAAttga ttgtaATGac tctatgtgca gtaccagtga cgacacggta tccgctactc
E4 orf start ->
               E4 cds ->
  3421 agattgttag acagctacaa cacgcctcca cgtcgacccc caaaaccgca tccgtgggca
  3481 ccccaaaacc ccacatccag acgccggcta ctaagcgacc tagacagtgt ggactcacag
  3541 agcagcacca cggacgtgtc aacacccacg tgcacaaccc gctcctgtgt tcaagtacaa
  3601 gtaacaacaa aagaaggaaa gtgtgtagtg gtaacactac gcctaTAAta cacttaaaag
                                                        <- E4 end
  3661 gtgacaaaaa cagtttgaaa tgtttaagat ataggctacg caaatatgca gaccattact
  3721 cagaaatatc ctccacctgg cattggacag gttgtaataa aaacactggt atattaactg
  3781 taacatataa tagtgaggta caaagaaata ccttttgga tgtagttact attcctaaca
```

```
3841 gtgtacaaat ctcggtggga tacatgacta taTGAatctg tatattgtat acagtatgta
                                       <- E2 end
                                E5 orf start ->
3901 acattactAT Gctatcttta gtgtttttat tgtgcttttc tgtgtgcctt tatgtgtgct
      E5 cds ->
3961 gcaatgtccc gcttgtgcag tctgtctatg tgtgtgcttt tgcttggttg ttggtgtttc
4021 tttttatagt tgttattaca tccccattaa cagcatttgc tgtatacatt tgttgctatt
4081 tactacctat gtttgtatta catatgcatg ctttacacac catacaaTAA ttactataat
                                                        <- E5 end
4141 gtacagtaca gtgtaacata cctgtgatgt gcatgttgtt gtatttttgt atttttgtat
4201 ttttgtattt ttgtatttta tatgtttaaT AAaccATGgt atcccaccgt gcagcacgtc
                                         -> ->
                               L2 orf start   L2 cds
4261 gcaagcgggc ctctgcaact gacttatata gaacatgtaa gcaatccggt acgtgccccc
4321 ctgatgttat taacaaagtg gaaggcacaa ccttagctga taaaatttta cagtggtcta
4381 gccttgggat atttttgggt ggccttggca ttggtACCGG CAGTGGTtct ggaggccgta
                                            -> E2 bind
4441 cgggctatgt acccttaggg ggcaggtcta atactgttgt ggatgttggc cccactaggc
4501 cacctgtggt tattgaacct gtagggccta ctgatccatc tattgttacg ttggtagagg
4561 attccagtgt tgttgcctct ggtgctccgg ttcccacatt taccggaacc tctgggtttg
4621 aaattacgtc ttctggtact accacACCAG CTGTGTTgga catcacacct accgtggact
                                 -> E2 bind
4681 ctgtttctat ttcgtcaact agttttacaa atcctgcatt ttctgatccc tctattattg
4741 aggtgcccca aacaggggag gtatcaggta atatatttgt tggtacacca acatcgggca
4801 gccatggata tgaggaaata cctttacaaa catttgcatc ttctgggtca ggtacggaac
4861 ccattagtag tacccccctc cctactgtgc ggcgggtacg gggtccccgc ctgtatagta
4921 gggctaatca acaggtccgt gtgtccacct cacagttttt aacacatccc tcatcgttgg
4981 ttacatttga taatccagct tatgagcccc tggacaccac actatccttt gagcctACCA
                                                          E2 bind ->
5041 GTAATGTTcc tgattccgat tttatggata ttattcgttt gcataggcca gcattatcct
5101 ctagacgtgg cactgttaga tttagtagat tgggtcaaag ggcaaccatg tttacacgta
5161 gtggtaaaca aatagggggt agggtacatt tttaccatga tataagcccc attgctgcta
5221 cagaggaaat tgaattgcag cctttaatta gtgctacaaa tgatagtgac ctgtttgatg
5281 tatatgcaga cttcccacct cctgcgtcca ctacacctag cactatacac aaatcattta
5341 catatccaaa gtattccttg accatgcctt ctactgctgc atcctcttac agtaatgtta
5401 cagtaccatt aacatctgca tgggatgtac ctatatatac tggcccggac attatattgc
5461 catcccatac tcctatgtgg cctagtacat ctcctaccaa tgcttccacc accACCTATA
                                                                -> E2 bind
5521 TAGGTattcA TGgcacacaa tattatttat ggccatggta ttattatttt cctaaaaaac
 L1 orf -> L1 ->
   start   cds
E2 bind <-
5581 gtaaacgtat tccctatttt tttgcagatg gctttgtggc ggccTAGtga cagtacggta
                                                    <- L2 end
5641 tatcttccac caccttctgt ggccagagtt gtcagcactg atgattatgt gtctcgcaca
5701 agcatatttt atcatgcagg cagttcccga ttattaactg taggcaatcc atattttagg
5761 gttgtaccta atggtgcagg taataaacag gctgttccta aggtatccgc atatcagtat
5821 agggtgttta gagtagcttt acccgatcct aataaatttg gattacctga ttctactata
5881 tataatcctg aaacacaacg tttggtttgg gcatgtgtag gtatggaaat tggtcgtggg
5941 cagcctttag gtattggcct aagtggccat ccattttata ataaattgga tgatacagaa
6001 agtgctcatg cagctacagc tgttattacg caggatgtta gggataatgt gtcagttgat
6061 tataagcaaa cacagctgtg tattttaggt tgtgtacctg ctattggtga gcactgggcc
6121 aagggcacac tttgtaaacc tgcacaattg caacctggtg actgtcctcc tttggaactt
6181 aaaaacacca ttattgagga tggtgatatg gtggatacag gttatggggc aatggatttt
6241 agtacattgc aggatacaaa gtgcgaggtt ccattagaca tttgtcaatc catctgtaaa
6301 tatccagatt atttgcaaat gtctgctgat ccctatgggg attctatgtt tttttgccta
6361 cgccgtgaac aactgtttgc aagacatttt tggaataggg caggtgttat gggtgacaca
6421 gtacctacgg acctatatat taaaggcact agcgctaata tgcgtgaaac ccctggcagt
6481 tgtgtgtatt ccccttctcc cagtggctct attattactt ctgattctca attatttaat
6541 aagccatatt ggttacataa ggcccagggc cataacaatg gtatttgttg gcataatcag
6601 ttgtttgtta ctgtagtgga cactacccgc agtactaatt taacattatg tgcctctaca
6661 caaaatcctg tgccaagtac atatgaccct actaagttta agcagtatag tagacatgtg
6721 gaggaatatg atttacagtt tatttttcag ttgtgcacta ttactttaac tgcagaggtt
6781 atgtcatata tccatagtat gaatagtagt atattagaaa attggaattt tggtgtccct
```

```
6841 ccaccaccta ctacaagttt ggtggataca tatcgttttg tgcaatcagt tgctgttacc
6901 tgtcaaaagg atactacacc tccagaaaag caggatccat atgataaatt aaagttttgg
6961 actgttgacc taaaggaaaa attttcctcc gatttggatc aatatcccct tggtcgaaag
7021 tttttagttc aggctgggtt acgtcgtagg cctaccatag gacctcgtaa gcgtcctgct
7081 gcttccacgt ctactgcatc tactgcatct aggcctgcca aacgtgtacg tatacgtagt
7141 aagaaaTAAt atgttagcac atatatgtat gtttgtatgt atggttttgt atgttgtatg
           <- L1 end
7201 tatgtatgta tttgtgtgat atattactgt attttgtttg tttgcgtgcg tgtatgtatg
7261 aatgtgcctt gtggcatgta tggtgttact gtacataatt gtggtattaa ataaagtatg
7321 ctaatagtgt tgtgtagggt tgcacccttg tgagtaacaa tactatttgt gtgtatgtgt
7381 attgctttgt accctatatt ctttcctgta tttcaagtta taaacttgca tactacacag
7441 catccatttt acttataatc ctccattttg ctgtgcaACC GATTTCGGTt gcctgtggct
                                              -> E2 bind
7501 tatatgtgac cttttaaaca taatacctaa actggcacat ttacaacccc tacatagttt
7561 aacctactgg cgcgccttct tggcgtacat gtggcacacc tggtattagt cattttcctg
7621 tccaggtgta ctaaaacaat ggcttgcaca actgtatcca caccctatgt aataaaactg
7681 cttttaggca catattttag tctgttttta cctgtgctaa ttgtataatt ggcgtgtaga
7741 accactttct tatccaacaa tctgtctact tgttacataa actataaact gactcactta
7801 tacatacata gtttatgcaA CCGAAAAAGG Ttgggcccta taacacatac cttttctt
                         -> E2 bind
```

```
LOCUS       HPV59MY911    452 bp ds-DNA             VRL       16-OCT-1994
DEFINITION  Human papillomavirus type 59 (HPV-59), partial L1 cds, My09/My11
            region.
ACCESSION   U12496
SOURCE      Human papillomavirus type 59 DNA.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Bernard,H.-U., Chan,S.-Y., Manos,M.M., Ong,C.-K., Villa,L.L.,
            Delius,H., Peyton,C.L., Bauer,H.M., and Wheeler,C.M.
  TITLE     Identification and assessment of known and novel human
            papillomaviruses by PCR amplification, restriction fragment
            length polymorphisms, nucleotide sequence, and phylogenetic
            algorithms
  JOURNAL   J. Infect. Dis. (1994) In press
COMMENT     HPV-59 was first isolated from a vulvar intraepithelial neoplasia
            of the genital mucosa.  Cloned HPV-59 DNA was obtained from the
            Papillomavirus Reference Center, Heidelberg and subsequently
            sequenced by Dr. H. Delius over the L1 MY09/MY11 segment.  HPV-59
            and the several other HPV types recently sequenced over this
            region by Dr. Delius were used as type-specific probes to screen
            DNA for novel genital HPV types.  The screened DNA was obtained
            from four recent epidemiological studies.  Primer regions are
            annotated in the sequence; information in this region is not
            accurate due to primer degeneracy.

BASE COUNT      130 a     88 c     83 g    151 t
ORIGIN
        1 gctcagggtt taaacaatgg tatatgttgg cacaatcaat tgtttttaac agttgtagat
  L1 cds ->
          -> MY11 PCR primer <-
       61 actactcgca gcaccaatct ttctgtgtgt gctctactac tctctattcc taatgtatac
      121 acacctacca gttttaaaga atatgccaga catgtggagg aatttgattt gcagtttata
      181 tttcaactgt gtaaaataac attaactaca gaggtaatgt catacattca taatatgaat
      241 accactattt tggaggattg gaattttggt gttacaccac ctcctactgc tagtttagtt
      301 gacacatacc gttttgttca atctgctgct gtaacttgtc aaaaggacac cgcaccgcca
      361 gttaaacagg acccttatga caaactaaag ttttggcctg tagatcttaa ggaaaggttt
      421 tctgcagatc ttgatcagtt tcctttggga cg
                                      L1 cds ->
                     -> MY09 PCR primer  <-
```

```
LOCUS       HPV68ME180   6042 bp ds-DNA             VRL       11-OCT-1991
DEFINITION  Human cellular DNA/Human papillomavirus type 68 proviral DNA.
ACCESSION   M73258
KEYWORDS    proviral gene.
SOURCE      Human papillomavirus DNA and Homo sapiens cervix DNA from an
            omental metastasis of a cervical carcinoma.
REFERENCE   1  (bases 1 to 6042)
  AUTHORS   Reuter,S., Delius,H., Kahn,T., Hofmann,B., Zur Hausen,H. and
            Schwarz,E.
  TITLE     Characterization of a novel human papillomavirus DNA in the
            cervical carcinoma cell line ME180
  JOURNAL   J. Virol. 65, 5564-5568 (1991)
COMMENT     The DNA of HPV-68/ME180 has been molecularly cloned from a genomic
            library of the cell line ME180 into bacteriophage lambda.  The cell
            line ME180 was established from an omental metastasis of a rapidly
            spreading cervical carcinoma and was assumed to harbour HPV-18 DNA.
            Under stringent conditions no hybridization was detected with a
            radiolabeled HPV-18 probe, but under reduced stringency three
            subgenomic fragments of HPV-18 hybridized to the HPV-68 DNA.  Reuter
            believes that other established human cervical carcinoma cell
            lines may harbor DNA of HPV types other than HPV-16 and HPV-18.

            The viral genome of HPV68 has been found integrated into the cellular
            genome of the ME180 cell line.  The integration of the genome requires
            linearization of the genome which most often occurs in the E1/E2
            region. This results in the inactivation of the E2 repressor of the
            E6 and E7 genes.  With repression removed, the E6 and E7 genes are
            able to immortalize keratinocytes.  Furthermore, the flanking cellular
            sequences may influence the expression of the E6 and E7 oncogenes.

            The 6042 bp sequence described in this entry contains both HPV-68
            DNA and cellular flanking DNA.  The HPV-68 sequence extends from bp
            1-5993, and the cellular DNA flanks the HPV-68 sequence to the end.
            The 5993 bp of HPV-68 DNA contains sequences that extend from
            within ORF E5 up to E1, include the upstream regulatory region
            (URR) and the complete ORFs L2, L1, E6, and E7.  Similar to HPV-16,
            the E1 ORF is disrupted by a frameshift mutation.  Another 897 bp
            segment, not included in this entry, had been cloned along with the
            included 6042 bp sequence.   This segment of HPV-68 DNA included
            the 5' end of the E5 ORF, the 3' end of E1, and the 5' end of E2.

BASE COUNT     1810 a   1155 c   1279 g   1798 t
ORIGIN
        1 tatgtatgtt gcactgtccc gcttctgcag tccatgcatg tgtgtgtgta tgtgtggata
       61 cttgtgtttg tgtttatatt agtacgtacc acaccattgg aggtctttgc tgtatatata
      121 ctttttttt tactgcctat gtgggtatta cacagttttg ctcgttatag tatgccttaa
      181 gttttgtatt gtgcatttgt attggtgtat atttttaTAA ataaatATGg tatcacaccg
                                        L2 orf start ->    -> L2 cds
      241 tgctgccagg cgcaagcgtg catctgcaac tgaattatat aaaacatgca aacaatcagg
      301 cacatgtcct cctgatgtta taaataaggt tgaaggcacc acacttgcag acaaactatt
      361 gcaatggacc agtttaggta ttttttgggg tggcctaggc attggtactg ggtcaggaac
      421 cgggggtcgt actgggtaca ttcctttagg tggtaaacct aatactgttg tagatgtttc
      481 gcctgcacgt ccacctgtgg ttattgaacc tgtgggtcct acagaaccct ccattgtgca
      541 attggtggaa gattccagtg ttattacatc tggcacaccg gtaccaacat ttacaggcac
      601 ttctgggttt gaaattacat cttcttctac cactacacct gctgtgttag acattacccc
      661 ttcgtctggg tctgtgcaag taagcagtac tagttttact aaccctgcat ttgcagaccc
      721 cactattata gaagtgcctc aaacaggtga agtctctggt aatgtgtttg taagtacccc
      781 cacatcggga acacatggat atgaagaaat acctatgcag gtatttgcaa cacatggcac
      841 tggtacagaa cctattagta gtacacctat acctggggtt agtcgtgtgg cagggccacg
      901 tttatatagt agggcacatc aacaggttcg tgttagtaat tttgattttg taactcaccc
      961 ttcatcattt gtaacatttg ataatcctgc ttttgagcct gttgatacta cacttacata
     1021 tgaacctgct gacatagctc ctgatccgga ttttctggac attgttcgtt tacataggcc
     1081 tgccttaact tcccgaagag gcacagtacg ttttagcaga gtaggcaaaa aggcaactat
     1141 gtttacacgc cggggtacac aaattggggc acaggtgcac tattatcatg atattagtgg
     1201 cattgctcct gctgacagca ttgaactaca acctttggtt gccccagagc agtctgaccc
     1261 tatggatact ttatatgata tatatgcacc agatactgac aatactacag tattggatac
```

```
1321 tgcattccat aatgctacat ttacctcccg ttcccatata tctgttcctt cattagcgtc
1381 tacagcatct actacatatg ctaacactac tattcctatt ggtactgctt ggaacacgcc
1441 tgtaaatact ggtcctgatg ttgtgttacc agcaacgtct ccacagttgc ctttaacacc
1501 ctctacacca attgatacaa cctatgccat aactatatat ggcaccaatt attatttatt
1561 accattattg ttctttttat TAAaaaaacg taaacgcctt ccttattttt ttgcagATGg
                L1 orf start ->                          L1 cds ->
1621 cattgtggcg ctcTAGcgac aacatggtgt atttgcctcc cccctcagtg gcgaaggttg
                   <- L2 end
1681 tcaatacaga tgattacgta acacgcactg gcatttatta ctatgctggt acatctaggt
1741 tattaactgt aggccatcca tattttaagg tccctatgtc tgggggccgc aagcaggaca
1801 ttcctaaggt gtctgcatat caatacaggg tgtttaggat ttccctacct gatcctaata
1861 aatttagtct tcctgagtct acattatata accctgatac gcagcgattg gtatgggcct
1921 gtgttggtgt tgaaataggt agggggcagc cattaggtgt tggccttagt gggcatccat
1981 tatataatag gctagatgat actgaaaatt ccccgttttc ctccaacaaa aatcctaagg
2041 acagtaggga caatgtttca gtggactata aacaaacgca actatgtatt ataggctgtg
2101 ttcctgccat tggggagcac tgggccaaag gtaaatcttg taagcctagc aatgtgcagc
2161 ccggggactg tccaccattg gaattagtaa atacacctat tcaggatggc gatatgattg
2221 atacaggata tggtgctatg gactttagta cattacaaga aacaaaaagc gaggtgcctt
2281 tagatatatg tcaatcagtc tgcaaatatc ctgactattt acaaatgtct gcagatgtat
2341 atggagacag tatgttcttt tgtttacgta gggaacagtt atttgctagg cattttgga
2401 atagaggggg catggtaggg gacactatac ctactgaatt gtatattaag ggcactgaca
2461 tacgtgacag tcctagtagt tatgtatatg ccccctcgcc tagtgggtct atggtatcct
2521 cagactccca gttatttaac aagccctatt ggctgcacaa ggcacaggga cacaacaatg
2581 gtatttgttg gcataatcaa ttatttctta ctgttgtgga taccactcgc agtaccaatt
2641 ttactttgtc tactactact gaatcagctg taccaaatat ttatgatcct aataaattta
2701 aggaatatat taggcatgtt gaggaatatg atttgcaatt tatatttcag ttgtgtacta
2761 taacattgtc cactgatgta atgtcctata tacatactat gaatcctgct attttggatg
2821 attggaattt tggtgttgcc cctccaccat ctgctagtct tgtagataca taccgctatc
2881 tgcaatcagc agcaattaca tgtcaaaaag acgcccctgc acctactaaa aaggatccat
2941 atgatggctt aaacttttgg aatgtaaatt taaaggaaaa gtttagttct gaactggacc
3001 agtttccttt aggacgcaaa tttcttttac aggcaggtgt ccgccgacga cccactatag
3061 gcccccgtaa acgccctgcc acagcaacta ctgcatctac ctctaagcac aaacgtaaac
3121 gtgtgtcaaa gTAAttgttg tatgttttgt tttgtatgtt ggttgtatgt gtggttgtat
                LCR ->
                  <- L1 end
3181 atgtgtcatg ttgttgttgg tatgttgtgc atgtatgtgt atatgtatat gtgtatgttt
3241 gcaggtatgt ttgtataatc tgtttttgtt AATAAAgtat gtatgtcagt ttactttgtg
                               signal ->
3301 gttgcaccct gtgactaaca tatgtccttg ttttacatat cataggactg caacatttcc
3361 tacataattt gtagccctac cctaaggtgt gttacagtac atgtaaTATA TATatagttc
                                                 signal ->
3421 tatattatac caagtggcca ttttgtaagg ccattttgtg tgcaaccgtt ttcggtcggt
3481 ggtgctattt ccttctatac agtattaaaa actatgtgtt tcagcaaaaa catgtttcac
3541 cttggtttac ccacatagtt ggcaccggta acagtatgta ctggcgcacc ttacttagtc
3601 atcatcctgt ccaggtgcag tgcaacaata gtttggcagc cTATATATct ccacccttgt
                                          signal ->
3661 AATAAAactg cttttaggca taggttttta actgttttta cttgcctaat agcatagttg
     -> signal
3721 gcctgtataa ctacttttgc attcaagaat gtgtcttgta gtgtaagtta tacagtgact
3781 aataccacat ccataaattt gtgcaaccga aataggttgg gcacacatac caatactttt
3841 acttataaca ttttacaatc attttatagt ataaagggag tgaccgaaaa cggtcatgac
3901 cgaaaacggt gTATATAAag cTGAacacag cagttgtcta taccaATGgc gctatttcac
          signal ->
                  E6 orf start ->              E6 cds ->
3961 aaccctgagg aacggccata caaattgcca gacctgtgca ggacattgga caccacattg
4021 catgacgtta caatagactg tgtctattgc agaaggcaac tacaacggac agaggtatat
4081 gaatttgcct ttggtgactt aaatgtagta tatagggacg gggtaccatt agctgcatgc
4141 caatcatgta ttaaatttta tgcgaaaata cgggaactac gatattactc agaatcggtg
4201 tatgcaacaa cattagaaac cataactaat acaaagttat atgatttatc aataaggtgc
4261 atgtgttgcc tgaaaccatt gagtcctgct gaaaaactaa ggcacctaaa ttcaaaacga
4321 agatttcata aaaTAGcagg aaactttaca ggacagtgtc gccactgctg gaccagtaaa
         E7 orf start ->
4381 cgagaggacc gcagacgcac acggcaggaa acacaagtaT AAactaactA TGcatggacc
                                                  <-E6 end -> E7 cds
```

```
4441 aaagcccacc gtgcaggaaa ttgtgttaga gttatgtcca tgcaatgaaa tagagccggt
4501 cgaccttgta tgtcacgagc aattaggaga ttcagacgat gaaatagatg aacccgacca
4561 tgcagttaat caccaccaac atcaactact agccagacgg gacgaacaac agcgtcacac
4621 aattcagtgt acgtgttgta agtgtaacaa cctactgcaa ctagtagtag aagcgtcgcg
4681 ggagaacctg cggaacgtag aactgctgtt tatggactca ctaaattttg tgtgtccgtg
4741 gtgtgcaacg gaaacccagT AAtctgcaAT Ggccaattgt gaaggtacag atggggacgg
              E1a orf start ->    -> E1a cds
                            <- E7 end
4801 gacggggtgt aacggatggt tttttgtaca agcaatagta gataaacaaa caggtgacac
4861 agtctcagag gatgaggatg aaaacgcgac agatacaggt tcagacatgg tagatttcat
4921 tgatgatgct acagatattt gtatacaggc agagcgTGAg acagcacagg tactgttaaa
                                  E1b orf start ->
4981 tatgcaacag gcccaaaggg atgcacaaac agtgcgtgcc ctaaaacgaa agtatacaga
5041 cagtatagaa agcagccctt tagcaaagtc gccattacag gaactatcaa tatggaagtg
5101 gaaactaact cggaggTAAc tgtagcaact aatacaaatg gggcggacgg ggaggatgaa
                       <- E1a end
5161 ggggaaaatg gcgacagcat acgggaggac tgtagtagtg tagacagtgc tatagatagt
5221 gaaaaccagg atcctaaatc acctactacg caactaaaag tattattaca atgtaataat
5281 aaaaaagctg caatgttaac agaatttaaa aaagtatatg gattgtcctt taatgaccta
5341 gtacgtacat ttaaaagtga taagaccaca tgtacggact gggtagcagc aatattcgga
5401 gtaaatccaa ccattgccga agggtttaaa acactaatta aacaatatgc attatatacc
5461 catatacaat gtttagatac aaaaaacgga atattaatat taatgttaat aagatacaaa
5521 tgtgggaaaa atagaataac agtaggaaaa ggattaagta cattgttgca tgttccagac
5581 agctgtatgc ttttgcagcc accaaaattg cgtagccctg ttgcagcatt gtattggtat
5641 agaacaggaa tatctaatat tagtgaggtg tgtggagaca cgccagaatg gataaaaaga
5701 ttaactataa tacaacatgg aatagatgat agtgtatttg atctatcaga catggtacaa
5761 tgggcatttg ataatgagtt aacagatgaa agtgatatag cattttcata tgctatgttg
5821 gcagattgta atagtaatgc tgcagcgttt ttaaaaagca actgtcaagc aaaatatgta
5881 aaagattgtg caacaatgtg tagacattac aaacgggcac aaaaacgaca aatgtcaatg
5941 ccgcaatgga ttaaatttag atgcagtaaa tgtgatgaag gcggtgattg gcgcatggac
                                                E1b partial <--> human
                                                                sequence
6001 tgcagaattt actacttgtg tttattacag gcgatttcct ga
```

LOCUS HPVCP141 455 bp ds-DNA VRL 16-OCT-1994
DEFINITION Human papillomavirus, isolate CP141, partial L1 cds, My09/My11 region.
ACCESSION U12476
SOURCE Human papillomavirus DNA derived from a cytologically normal cervical sample from a Hispanic woman, 20 years of age, isolate CP141.
REFERENCE 1 (bases 1 to 455)
AUTHORS Peyton,C.L. and Wheeler,C.M.
TITLE Identification of five novel human papillomaviruses in the New Mexico triethnic population
JOURNAL J. Infect. Dis. (1994) In press
COMMENT Data kindly provided prior to publication by Dr. C. Wheeler, University of New Mexico, School of Medicine, New Mexico Tumor Registry, 900 Camino del Salad NE, Albuquerque, NM, 87131-5306.

Five novel HPV sequences were identified in a study in which 3655 cervical specimens were screened against known genital HPV DNA [1]. The specimens were obtained from clinical investigations conducted at the University of New Mexico. The study subjects included Native Indians, Hispanics, and non-Hispanic whites. CP141 was derived from a cytologically normal cervical sample from a Hispanic woman, 20 years of age. The viral DNA was PCR amplified using the L1 consensus primer MY09/MY11 pair, which can hybridize to a broad spectrum of HPV types. Resultant fragments range from 449 to 458 nucleotides in length. The amplification products were initially screened against 2 sets of type-specific probes and a generic probe. If hybridization to the generic probe and not to the type-specific probes ocurred, the samples were further analyzed by restriction fragment length polymorphisms. RFLP patterns which did not match reference patterns were considered to be derived from novel HPVs. The five novel samples which were were identified in this study include CP8304, CP6108, CP8061, CP141, CP4173. Peyton et al. also identified two HPV45 subtypes and one HPV56 subtype. They conclude that since the existence of subtypes appears to be relatively rare, it suggests that HPV45 and HPV56 are more divergent than many HPV types. It should be noted that CP141 (U12476) is almost identical to LVX160 (U12486) and HPVL1AE1 (U01535) and that CP4173 (U12477) is almost identical to LVX100 (U12485). Both LVX160 and LVX100 were identified by Ong et al. in a 1994 study which examined Amazonian Indian subjects (Ong et al., J. Infect. Dis., 1994, in press). Primer regions are annotated in the sequence; information in this region is not accurate due to primer degeneracy.

In a subsequent study Bernard et al. evaluated ten novel genital HPV types, including the five identified in the Peyton et al. study, and other known genital types to determine phylogenetic relationships. They observed that the genital types CP6108, CP8304, CP4173 and CP8061 form a branch with HPV types 61 and 62. This emergent minor branch is positioned between two others which contain cutaneous types. Bernard et al. speculate as to whether other low-risk genital types have escaped detection because of considerable sequence divergence from the common genital types (Bernard et al., J. Infect. Dis., 1994, in press).

Bernard et al. also assessed the linear correlation coefficients for the MY9/MY11 fragments against the rest of L1 (.851) and against the E6 gene (.888). Since these values are close, the authors suggest that the evolutionary distance information obtained for the primer pair region should be comparable to that available from the other regions of the genome (Bernard et al., J. Infect. Dis., 1994, in press).

BASE COUNT 142 a 86 c 90 g 137 t

```
ORIGIN
        1 gcccagggaa ctaataatgg catttgttgg cataaccagt tgtttattac tgtggtggac
  L1 cds ->
          -> MY11 PCR primer <-
       61 actacacgta gtactaattt tacattgtct gcctgcaccg aaacggccat acctgctgta
      121 tatagcccta caaagtttaa ggaatatact aggcatgtgg aggaatatga tttacaattt
      181 atatttcaat tgtgtactat cacattaact gcagacgtta tggcctacat ccatactatg
      241 aatcctgcaa ttttggacaa ttggaatata ggagttaccc ctccaccatc tgcaagcttg
      301 gtggacacgt ataggtattt acaatcagca gctatagtat gtcaaaagga tgctcctaca
      361 cctgaaaaaa aggatcccta tgacgattta aaattttgga atgttgattt aaaggaaaag
      421 tttagtacag aactagatca gtttcctctg ggacg
                                          L1 cds ->
                          -> MY09 PCR primer  <-
```

HPVL1AE1

```
LOCUS       HPVL1AE1      415 bp    DNA             VRL       25-MAY-1994
DEFINITION  Human papillomavirus, L1 partial cds, My09/My11 region.
ACCESSION   U01535
SOURCE      Human papillomavirus DNA, PCR amplified clone AE1
REFERENCE   1  (bases 1 to 415)
  AUTHORS   Tachezy,R., Van Ranst,M.A., Cruz,Y. and Burk,R.D.
  TITLE     Consensus primer mediated PCR allows identification of novel human
            papillomavirus PCR-types in cervicovaginal lavages
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 415)
  AUTHORS   Van Ranst,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-1993) Marc A. Van Ranst, Albert Einstein College
            of Medicine, Dept. of Microbiology & Immunology, 1300 Morris Park
            Avenue, Bronx, NY 10461, USA
COMMENT     Isolate AE1 is a novel HPV type identified in a study conducted to
            screen cervical lavages from over 500 women for novel HPV types.
            All women involved in the study were seen by physicians from
            clinics or private practices in the Bronx, N.Y. area.  Derived
            sequences are PCR products amplified over the My09/My11 primer
            region of L1.  The sequence for AE1 is virtually identical to those
            for HPVCP141 (U12476) and HPVLVX160 (U12486).
BASE COUNT      132 a     78 c     78 g    127 t
ORIGIN
        1 catttgttgg cataaccagt tgtttattac tgtggtggac actacacgta gtactaattt
   L1 cds ->
       61 tacattgtct gcctgcaccg aaacggccat acctgctgta tatagcccta caaagtttaa
      121 ggaatatact aggcatgtgg aggaatatga tttacaattt atatttcaat tgtgtactat
      181 cacattaact gcagacgtta tggcctacat ccatactatg aatcctgcaa ttttggacaa
      241 ttggaatata ggagttaccc ctccaccatc tgcaagcttg gtggacacgt ataggtattt
      301 acaatcagca gctatagcat gtcaaaagga tgctcctaca cctgaaaaaa aggatcccta
      361 tgacgattta aaattttgga atgttgattt aaaggaaaag tttagtacag aacta
                                                              L1 cds ->
```

```
LOCUS       HPVLVX160     455 bp ds-DNA             VRL       16-OCT-1994
DEFINITION  Human papillomavirus, isolate LVX160, partial L1 cds, My09/My11
            region.
ACCESSION   U12486
SOURCE      Human papillomavirus, isolate LVX160, from cervical smear.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   Ong,C.-K., Bernard,H.-U. and Villa,L.L.
  TITLE     Identification of genomic sequences of three novel human
            papillomaviruses in cervical smears of Amazonian Indians
  JOURNAL   J. Infect. Dis.  (1994) In press
COMMENT     HPVLVX82, HPVLVX100 and HPVLVX160 were found in cervical smears
            taken from members of isolated Amazonian tribes. The samples were
            PCR-amplified using the MY09/My11 consensus primers, then examined
            in hybridization experiments in order to determine their homology
            with known HPV types. Each of these three novel variants were more
            than 10% divergent from their closest known relatives, suggesting
            that they may qualify to be considered new types. Although the
            tribes were thought to have been sexually isolated from
            non-Amerindian populations for at least 12,000 years, sequences
            closely related to these novel variants have since been detected in
            other distinct populations.  Ong et al. believe this may be
            evidence for the hypothesis that papillomavirus types evolved
            before the speciation of Homo sapiens, and consequently before the
            divergence of ethnic groups. LVX160 is virtually identical to
            HPVL1AE1 (U01535) and to HPVCP141 (U12476).  Primer regions are
            annotated in the sequence; information in this region is not
            accurate due to primer degeneracy.  All similarity calculations
            exclude data from this region.

BASE COUNT      142 a     86 c     90 g    137 t
ORIGIN
        1 gcacagggtc ataataatgg catttgttgg cataaccagt tgtttattac tgtggtggac
  L1 cds ->
          -> MY11 PCR primer <-
       61 actacacgta gtactaattt tacattgtct gcctgcaccg aaacggccat acctgctgta
      121 tatagcccta caaagtttaa ggaatatact aggcatgtgg aggaatatga tttacaattt
      181 atatttcaat tgtgtactat cacattaact gcagacgtta tggcctacat ccatactatg
      241 aatcctgcaa ttttggacaa ttggaatata ggagttaccc ctccaccatc tgcaagcttg
      301 gtggacacgt ataggtattt acaatcagca gctatagcat gtcaaaagga tgctcctaca
      361 cctgaaaaaa aggatcccta tgacgattta aaattttgga atgttgattt aaaggaaaag
      421 tttagtacag aactagatca gtttcctctg ggacg
                                       L1 cds ->
                          -> MY09 PCR primer  <-
```