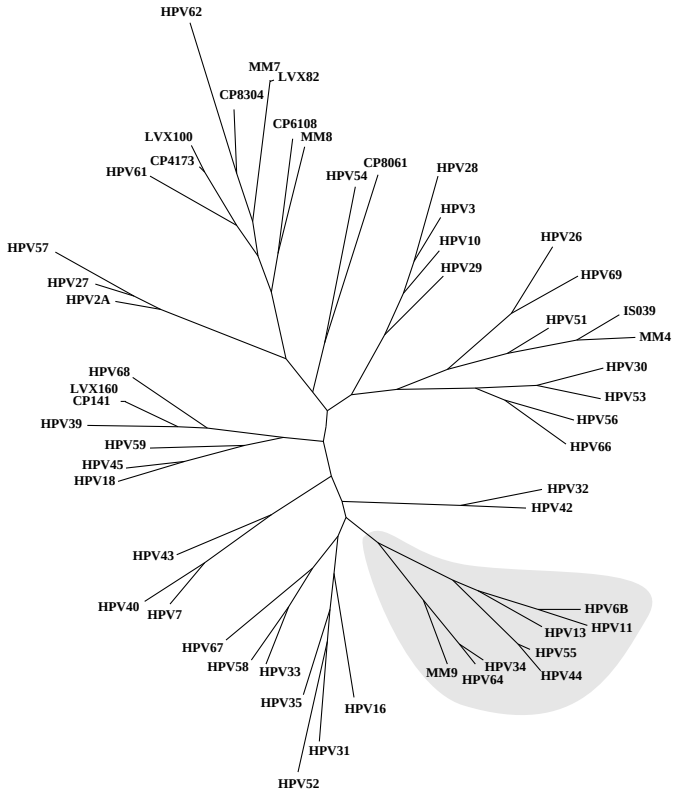
Group B Sequences

HPV6b	**HPV11**
HPV13	**HPV34**
HPV44	**HPV55**
HPV64	**HPVMM9**



INTRODUCTION

Group B consists of human papillomavirus types 6b, 11, 13, 34, 44, 55, 64, and the novel virus MM9, a group primarily associated with orogenital lesions with low oncogenic potential. Lorinz et al. classified HPV-6, HPV-11 and HPV-44 as "low risk" viruses [1]. DNA from these three viruses and others in the low-risk class was detected in 20.2% of the low-grade cervical lesions, in 4.2% of the high-grade lesions, and in none of the 153 invasive cancers screened [1].

Many researchers view HPV-6 and HPV-11 together as a functional group [1]. These two viruses are primarily responsible for the benign HPV infection of the anogenital tract. Condylomata acuminata have been shown to harbour HPV-6 or HPV-11 DNA in more than 93% of the cases [1]. Conversely, relatively few HPV-6 and HPV-11 positive genital malignancies have been identified despite extensive international screening. One type of malignancy, although rare, is strongly correlated with HPV-6 and HPV-11 infection: Buschke-Lowenstein tumors, the highly differentiated squamous cell tumors of the genital region, are associated almost exclusively with HPV-6 and HPV-11 [2,3]. HPV-44 and HPV-55, also group B viruses, have been detected in condyloma acuminata of the genital region (vulvar and penile, respectively) [4,5].

The strong association of HPV-6 and HPV-11 with certain types of genital carcinomas appears to be inconsistent with their classification as low-risk. Several explanations have been proposed to explain this anomaly. First, Lorincz et al. suggest the limited number of papillomavirus probes available to researchers may have contributed to the false-positive identification of HPV-6 and HPV-11 [1]. Second, researchers have shown a correlation with oncogenic potential and the presence of a duplicate upstream regulatory region in the genome. A species of HPV-11 with this duplication has been shown to transform baby rat kidney cells and such duplications have been found in carcinomas harbouring HPV-11 and HPV-6 DNA [6–9]. Rubben et al. suggest that cellular and environmental factors following infection may induce this duplication event and/or other rearrangements leading to acquired oncogenic properties [9]. Cofactors which may contribute to malignancy include alcohol and tobacco use and sexual intercourse during menstrual periods [10, 11].

In addition to their involvement with anogenital tract lesions, HPV-6, HPV-11 and HPV-13 are strongly associated with ororespiratory tract infection. In one study, 72% of all laryngeal papillomas and 25% of all oral papillomas were positive for HPV-6 and HPV-11 DNA [12]. HPV6 and HPV-11 have been detected in benign papillomas infecting almost every epithelial lining of the upper digestive and respiratory tracts. These tissues include the larynx, sinonasal area, lung, tonsil, tongue, and linings of the oral cavity [6, 12–14]. Unexpectedly, a high percentage (60%) of laryngeal carcinomas have been shown to be positive for HPV-11 DNA [15]. HPV-13, another HPV type highly correlated with oral lesions, was reported by de Villiers to be present in 13% of all oral papillomas [12]. Specifically, HPV-13 has been highly correlated with oral focal epithelial hyperplasia (FEH), a benign lesion situated primarily on the mucosae of the lower lips and cheeks [16]. This disease is frequently found among Indians in Central and South America and in Eskimos

in Greenland and Alaska [16]. However, the prevalence among Caucasians in the same area is much lower [16]. HPV-13 has also been detected in a case of low-grade cervical dysplasia and in Bowenoid papulosis in an HIV-positive male [16].

Several subtypes of HPV-6 have been identified. Subtype 6a has been isolated from tonsillar carcinoma, lung carcinoma and Buschke-Lowenstein tumors [8, 13, 17]. HPV-6b, the prototypical HPV-6 subtype, was initially cloned and sequenced from a benign genital wart [18]. It has been subsequently detected in various genital and upper digestive and respiratory tract lesions. The HPV-6c genome was molecularly cloned from both a respiratory-tract papilloma and a condyloma acuminatum of the cervix [19]. This subtype has also been detected in benign laryngeal papillomas and benign nasopapillomas [20]. The HPV-6d genome, cloned from Buschke-Lowenstein tumors, contains a tandem duplication of 459 base pairs in the noncoding region of the genome [13]. HPV-6e was identified in a genital wart and laryngeal papillomas [19–21]. HPV-6f has been cloned from a benign laryngeal papilloma and a non-inverted nasal papilloma [20, 22]. And, finally, HPV-6vc was cloned from a rapidly growing vulvar verrucous carcinoma [23].

The other viruses in this group, HPV-34, HPV-64, and the novel virus MM9, have been predominantly linked to anogenital lesions. HPV-34 was initially isolated and cloned from a squamous cell carcinoma of Bowen's type and subsequently detected in a genital intraepithelial neoplasia and periungual Bowen's disease [24]. A study which probed lesions with Bowen's disease and squamous cell carcinomas for HPV-34 DNA, reported only one case of positive hybridization, indicating that HPV-34 infection of this nature is relatively rare [24]. HPV-64, a recently identified virus, was cloned and isolated from a vulvar intraepithelial neoplasia [25]. MM9 was derived from a genital swab specimen. Initial prevalence data for MM9 is similar to that obtained for characterized "intermediate-risk" viruses [26].

Of the members of Group B, complete genomic sequences are available for HPV-6b, HPV-11, HPV-13 and HPV-34. HPV-44 has been sequenced over all of E6 and over the My09-My11 fragment of L1, while HPV-55, HPV-64 and the novel variant HPVMM9 have been sequenced only over the latter region. We consider HPV-11 and HPV-6b to be "close types"—sequences which qualify to be distinct types under the criterion of ten percent dissimilarity at the nucleotide level, but between which most of the changes are "silent", causing no difference at the amino acid level (Part III). In addition, the Pygmy Chimpanzee papillomavirus (PCPV-1) is close enough to HPV-13 for these to be considered "close types", although for the purposes of presentation, we have included PCPV-1 in Group I with the other animal papillomaviruses. Although HPV-34 was originally included in Group B on the basis of its association with the rest of the members of this group in the phylogenetic tree used to define the groups, it has since proven to be somewhat problematic in terms of its classification. It seems to cluster with different groups in different analyses, although its association with any group is generally rather weak. For this reason, we chose to exclude HPV-34 from Group B in the context of those analyses that were done "by group". In future it may be prove to be more reasonable to reclassify HPV-34 as a member of Group F, the "catch-all" group; however, we have elected to wait for further evidence before this decision is made. If this turns out to be the case, HPV-64 and HPVMM9 may have to be regrouped as well, since these cluster with HPV-34 on the "group-defining" tree.

---

[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] Gissmann,L., deVilliers,E.M., and zur Hausen, H. Analysis of human genital warts (condylomata acuminata) and other genital tumors for human papillomavirus type 6 DNA. *Int J Cancer* **29**:143–6 (1982)

[3] Lehn, H., Ernst,T.M., and Sauer,G. Transcription of episomal papillomavirus DNA in human condylomata acuminata and Buschke-Lowenstein tumours. *J Gen Virol* **65**:2003–10 (1984)

[4] Lorincz,A.T., Quinn,A.P., Goldsborough,M.D., Schmidt,B.J., and Temple, G.F. Cloning and partial DNA sequencing of two new human papillomavirus types associated with condylomas and low-grade cervical neoplasia. *J Virol* **63**: 2829–34 (1989)

[5] Favre,M., Kremsdorf,D., Jablonska,S., Obalek,S., Pehau-Arnaudet,G., Croissant, O., and Orth,G. Two new human papillomavirus types (HPV54 and 55) characterized from genital tumours illustrate the plurality of genital HPVs. *Int J Cancer* **45**:40–6 (1990)

[6] DiLorenzo,T.P., Tamsen,A., Abramson,A.L., and Steinberg,B.M. Human papillomavirus type 6a DNA in the lung carcinoma of a patient with recurrent laryngeal papillomatosis is characterized by a partial duplication. *J Gen Virol* **73**:423–8 (1992)

[7] Rosen,M. and Auborn,K. Duplication of the upstream regulatory sequences increases the transformation potential of human papillomavirus type 11. *Virology* **185**:484–7 (1991)

[8] Boshart,M. and zur Hausen,H. Human papillomaviruses in Buschke-Lowenstein tumors: physical state of the DNA and identification of a tandem duplication in the noncoding region of a human papillomavirus 6 subtype. *J Virol* **58**:963–6 (1986)

[9] Rubben,A., Beaudenon,S., Favre,M., Schmitz,W., Spelten,B., and Grussendorf-Conen,E.I. Rearrangements of the upstream regulatory region of human papillomavirus type 6 can be found in both Buschke-Lowenstein tumours and in condylomata acuminata. *J Gen Virol* **73**:3147–53 (1992)

[10] Rohan,T., Mann,V., McLaughlin,J., Harnish,D.G., Yu,H., Smith,D., Davis,R., Shier,R.M., and Rawls,W. PCR-detected genital papillomavirus infection: prevalence and association with risk factors for cervical cancer. *Int J Cancer* **49**:856–60 (1991)

[11] 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)

[12] de Villiers,E.M., Weidauer,H., Le,J.Y., Neumann,C., and zur Hausen, H. Papilloma viruses in benign and malignant tumors of the mouth and upper respiratory tract. *Laryngol Rhinol Otol (Stuttg)* **65**:177–9 (1986)

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

[14] Naghashfar,Z., Sawada,E., Kutcher,M.J., Swancar,J., Gupta,J., Daniel, R., Kashima,H., Woodruff, J.D., and Shah,K. Identification of genital tract papillomaviruses HPV-6 and HPV-16 in warts of the oral cavity. *J Med Virol* **17**:313–24 (1985)

[15] Syrjanen,S., Syrjanen,K., Mantyjarvi,R., Collan,Y., and Karja,J. Human papillomavirus DNA in squamous cell carcinomas of the larynx demonstrated by in situ DNA hybridization. *ORL J Otorhinolaryngol Relat Spec* **49**:175–86 (1987)

[16] Van Ranst,M., Fuse,A., Fiten,P., Beuken,E., Pfister,H., Burk,R.D., and Opdenakker,G. Human papillomavirus type 13 and pygmy chimpanzee papillomavirus type 1: comparison of the genome organizations. *Virology* **190**:587–96 (1992)

[17] Kahn.T., Turazza,E., Ojeda,R., Bercovich,A., Stremlau,A., Lichter,P., Poustka,A., Grinstein,S., and zur Hausen,H. Integration of human papillomavirus type 6a DNA in a tonsillar carcinoma: chromosomal localization and nucleotide sequence of the genomic target region. *Cancer Res* **54**:1305–12 (1994)

[18] de Villiers,E.M., Gissmann,L., and zur Hausen,H. Molecular cloning of viral DNA from human genital warts. *J Virol* **40**:932–935 (1981)

[19] Metcalfe,L., Chen,S.L., and Mounts,P. Structural analysis of human papillomavirus type 6c isolates from condyloma acuminatum and juvenile-onset and adult-onset laryngeal papillomata. *Virus Genes* **3**:11–27 (1989)

[20] Ward,P. and Mounts,P. Heterogeneity in mRNA of human papillomavirus type-6 subtypes in respiratory tract lesions. *Virology* **168**:1–12 (1989)

[21] Ishibashi,T., Tsunokawa,Y., Nomura,Y., Miyata,M., Terada,M. Molecular cloning and characterization of a new subtype of human papillomavirus 6 DNA. *Eur Arch Otorhinolaryngol* **248**:205–8 (1991)

[22] Ogura,H., Kawakami,T., Fujiwara,T., Sakai,A., Saito,R., Watanabe,S., Masuda,Y., and Yabe,Y. Detection of human papillomavirus type 6f genome in nasal papillomatosis. *Acta Otolaryngol (Stockh)* **112**:115–9 (1992)

[23] Rando,R.F., Sedlacek,T.V., Hunt,J., Jenson,A.B., Kurman,R.J., and Lancaster,W.D. Verrucous carcinoma of the vulva associated with an unusual type 6 human papillomavirus. *Obstet Gynecol* **67**:70S–75S (1986)

[24] Kawashima,M., Jablonska,S., Favre,M., Obalek,S., Croissant,O., and Orth,G. Characterization of a new type of human papillomavirus found in a lesion of Bowen's disease of the skin. *J Virol* **57**:688–92 (1986)

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

[26] Manos,M.M., Waldman,J., Zhang,T., Greer,C., Eichinger,G., Schiffmann,M., and Wheeler, C. Epidemiology and partial nucleotide sequence of four novel genital human papillomaviruses. *J. Infect. Dis.* (1994) In press

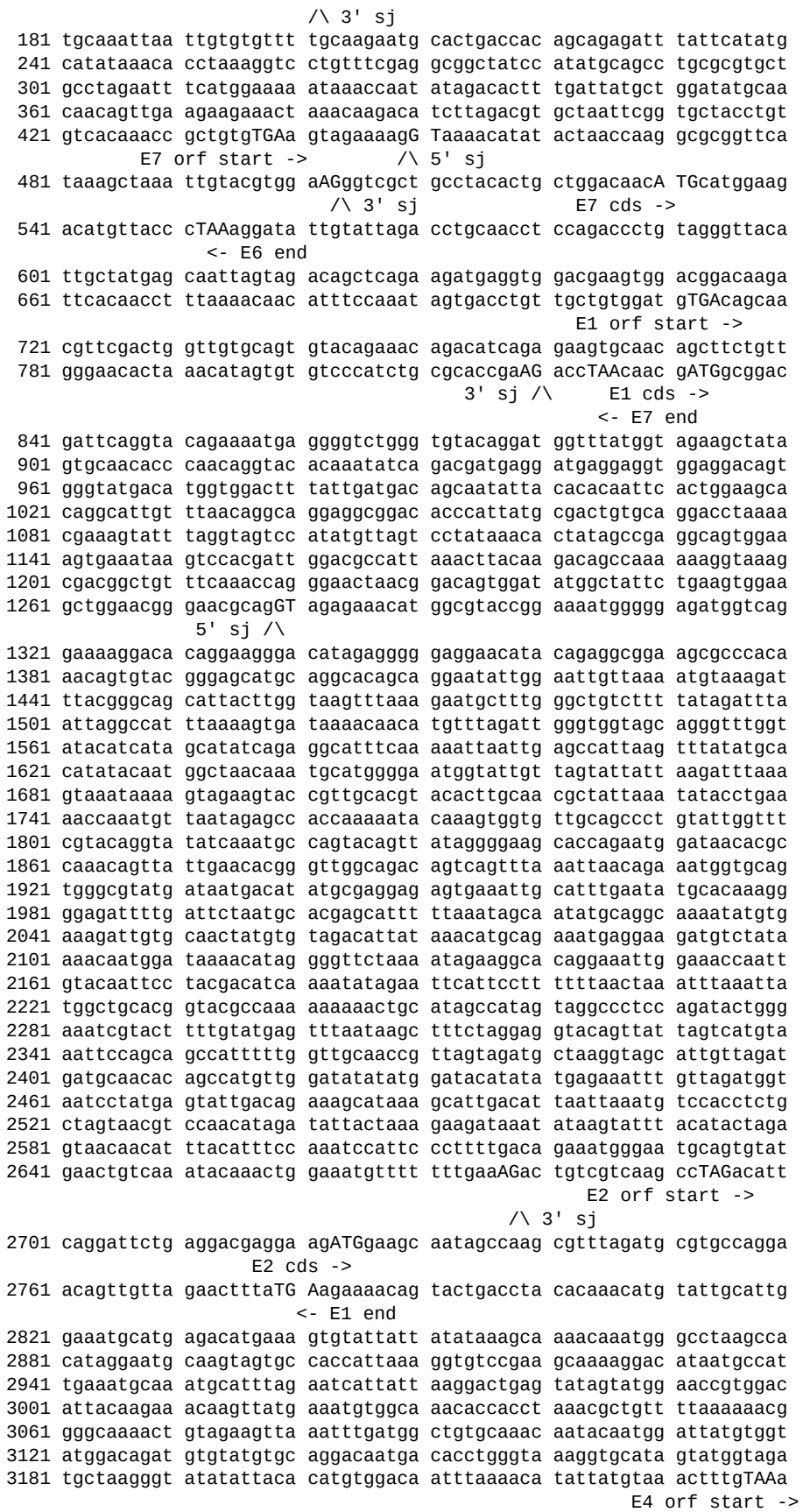
```
LOCUS       HPV6b        7902 bp ds-DNA   circular  VRL       11-MAR-1994
DEFINITION  Human papillomavirus type 6b (HPV-6b), complete genome.
ACCESSION   X00203
KEYWORDS    complete genome; overlapping genes.
SOURCE      Human papilloma virus type 6b DNA.
REFERENCE   1  (bases 1 to 7902)
  AUTHORS   Schwarz,E., Duerst,M., Demankowski,C., Lattermann,O., Zech,R.,
            Wolfsperger,E., Suhai,S. and Zur Hausen,H.
  TITLE     DNA sequence and genome organization of genital human
            papillomavirus type 6b
  JOURNAL   EMBO J. 2, 2341-2348 (1983)
COMMENT     HPV-11 and HPV-6 are responsible for the large majority of
            exophytic condylomas in the  genital tract.  Even though these
            lesions are frequently present in the genital tract, they are
            virtually absent in higher grade neoplasias and in cervical
            cancers.  HPV-6  also infects other mucosal types; the respiratory
            tract, oral cavity and conjunctiva.  It has been recovered from
            approximately 50% of respiratory tract lesions and 50% of all
            childhood conjunctival papillomas.  Respiratory papillomatosis is
            a rare disease that can be life-threatening because of its
            recurrent nature and the possibility of obstruction of the airways
            and respiratory distress.  The most frequent sites of infection are
            the vocal cords in the larynx, but papillomas may also be present
            in the trachea, lungs, nose and oral cavity.  These respiratory
            papillomas progress to malignancy rarely, as they account for less
            than 0.1% of all respiratory cancers.

            The 7902 bp complete genome of HPV-6b has been cloned in pBR322 and
            in lambda and was originally recovered from a genital wart.  The
            sense strand has been numbered by comparative analysis with BPV-1
            and HPV-1a.  Both the E6 and E7 ORFs contain conserved Cys-X-X-Cys
            cysteine doublet motifs.  The E6 ORF contains four of these motifs
            separated by 29, 36 and 29 intervening amino acids, while the E7
            ORF contains just two separated by 29 amino acids.  The E6 ORF also
            contains a small intron.  The E5a ORF codes for a protein of 91
            amino acids.  It has a stretch of 13 amino acids which is very rich
            in leucine.  The L2 ORF contains an extremely conserved cluster of
            basic residues both at the N terminus and at the C terminus ends.
            The authors feel that the conserved region of this part of this
            peptide may interact with the conserved L1 structural peptide,
            where the variable region may be involved with host or tissue
            specific functions.

            Between the end of L1 and the beginning of E6 lies a small open
            reading frame E8.  The first methionine is located in the middle of
            the ORF and it has no analog to other papillomaviruses sequenced at
            the time of publication.  In light of these facts, this ORF is
            probably not functional.  Thus, the region from the end of L1 to
            the beginning of E6 is probably the noncoding region containing the
            promoter and origin of replication.  Within the first segment of
            this region lies a monotonous repetition of thymine-purine which is
            just slightly disturbed.  Two repeats can be identified within the
            LCR; a 24 bp tandem repeat and a 9 bp direct repeat.  A TATA box is
            located at position 64 and a cap site is located directly in front
            of the E6 methionine codon.

BASE COUNT     2438 a   1530 c   1699 g   2235 t
ORIGIN
        1 GTTAATAACA ATCttggttt aaaaaaTAGg agggaccgaa aacggttcaa ccgaaaacgg
             <- E8 end      E6 orf start ->
       ->  CAAT-box  <-
     (begins at bp 7898)
       61 ttgTATATAA accagcccta aaatttagca aacgaggCAT TATGgaaagt gcaaatgcct
             -> signal                          E6 cds ->
                                       cap site ->   <-
      121 ccacgtctgc aacgaccatA Gaccagttgt gcaagacgtt taatctatct atgcatacgt
```

HPV6b


/\ 3' sj
181 tgcaaattaa ttgtgtgttt tgcaagaatg cactgaccac agcagagatt tattcatatg
241 catataaaca cctaaaggtc ctgtttcgag gcggctatcc atatgcagcc tgcgcgtgct
301 gcctagaatt tcatggaaaa ataaaccaat atagacactt tgattatgct ggatatgcaa
361 caacagttga agaagaaact aaacaagaca tcttagacgt gctaattcgg tgctacctgt
421 gtcacaaacc gctgtgTGAa gtagaaaagG Taaaacatat actaaccaag gcgcggttca
E7 orf start -> /\ 5' sj
481 taaagctaaa ttgtacgtgg aAGggtcgct gcctacactg ctggacaacA TGcatggaag
/\ 3' sj E7 cds ->
541 acatgttacc cTAAaggata ttgtattaga cctgcaacct ccagaccctg tagggttaca
<- E6 end
601 ttgctatgag caattagtag acagctcaga agatgaggtg gacgaagtgg acggacaaga
661 ttcacaacct ttaaaacaac atttccaaat agtgacctgt tgctgtggat gTGAcagcaa
E1 orf start ->
721 cgttcgactg gttgtgcagt gtacagaaac agacatcaga gaagtgcaac agcttctgtt
781 gggaacacta aacatagtgt gtcccatctg cgcaccgaAG accTAAcaac gATGgcggac
3' sj /\ E1 cds ->
<- E7 end
841 gattcaggta cagaaaatga ggggtctggg tgtacaggat ggtttatggt agaagctata
901 gtgcaacacc caacaggtac acaaatatca gacgatgagg atgaggaggt ggaggacagt
961 gggtatgaca tggtggactt tattgatgac agcaatatta cacacaattc actggaagca
1021 caggcattgt ttaacaggca ggaggcggac acccattatg cgactgtgca ggacctaaaa
1081 cgaaagtatt taggtagtcc atatgttagt cctataaaca ctatagccga ggcagtggaa
1141 agtgaaataa gtccacgatt ggacgccatt aaacttacaa gacagccaaa aaaggtaaag
1201 cgacggctgt ttcaaaccag ggaactaacg gacagtggat atggctattc tgaagtggaa
1261 gctggaacgg gaacgcagGT agagaaacat ggcgtaccgg aaaatggggg agatggtcag
5' sj /\
1321 gaaaaggaca caggaaggga catagagggg gaggaacata cagaggcgga agcgcccaca
1381 aacagtgtac gggagcatgc aggcacagca ggaatattgg aattgttaaa atgtaaagat
1441 ttacgggcag cattacttgg taagtttaaa gaatgctttg ggctgtcttt tatagattta
1501 attaggccat ttaaaagtga taaaacaaca tgtttagatt gggtggtagc agggtttggt
1561 atacatcata gcatatcaga ggcatttcaa aaattaattg agccattaag tttatatgca
1621 catatacaat ggctaacaaa tgcatgggga atggtattgt tagtattatt aagatttaaa
1681 gtaaataaaa gtagaagtac cgttgcacgt acacttgcaa cgctattaaa tatacctgaa
1741 aaccaaatgt taatagagcc accaaaaata caaagtggtg ttgcagccct gtattggttt
1801 cgtacaggta tatcaaatgc cagtacagtt ataggggaag caccagaatg gataacacgc
1861 caaacagtta ttgaacacgg gttggcagac agtcagttta aattaacaga aatggtgcag
1921 tgggcgtatg ataatgacat atgcgaggag agtgaaattg catttgaata tgcacaaagg
1981 ggagattttg attctaatgc acgagcattt ttaaatagca atatgcaggc aaaatatgtg
2041 aaagattgtg caactatgtg tagacattat aaacatgcag aaatgaggaa gatgtctata
2101 aaacaatgga taaaacatag gggttctaaa atagaaggca caggaaattg gaaaccaatt
2161 gtacaattcc tacgacatca aaatatagaa ttcattcctt ttttaactaa atttaaatta
2221 tggctgcacg gtacgccaaa aaaaaactgc atagccatag taggccctcc agatactggg
2281 aaatcgtact tttgtatgag tttaataagc tttctaggag gtacagttat tagtcatgta
2341 aattccagca gccatttttg gttgcaaccg ttagtagatg ctaaggtagc attgttagat
2401 gatgcaacac agccatgttg gatatatatg gatacatata tgagaaattt gttagatggt
2461 aatcctatga gtattgacag aaagcataaa gcattgacat taattaaatg tccacctctg
2521 ctagtaacgt ccaacataga tattactaaa gaagataaat ataagtattt acatactaga
2581 gtaacaacat ttacatttcc aaatccattc ccttttgaca gaaatgggaa tgcagtgtat
2641 gaactgtcaa atacaaactg gaaatgtttt tttgaaAGac tgtcgtcaag ccTAGacatt
E2 orf start ->
/\ 3' sj
2701 caggattctg aggacgagga agATGgaagc aatagccaag cgtttagatg cgtgccagga
E2 cds ->
2761 acagttgtta gaactttaTG Aagaaaacag tactgaccta cacaaacatg tattgcattg
<- E1 end
2821 gaaatgcatg agacatgaaa gtgtattatt atataaagca aaacaaatgg gcctaagcca
2881 cataggaatg caagtagtgc caccattaaa ggtgtccgaa gcaaaaggac ataatgccat
2941 tgaaatgcaa atgcatttag aatcattatt aaggactgag tatagtatgg aaccgtggac
3001 attacaagaa acaagttatg aaatgtggca aacaccacct aaacgctgtt ttaaaaaacg
3061 gggcaaaact gtagaagtta aatttgatgg ctgtgcaaac aatacaatgg attatgtggt
3121 atggacagat gtgtatgtgc aggacaatga cacctgggta aaggtgcata gtatggtaga
3181 tgctaagggt atatattaca catgtggaca atttaaaaca tattatgtaa actttgTAAa
E4 orf start ->

```
3241 AGaggcagaa aagtATGgga gcaccaaaca ttgggaagta tgttatggca gcacagttat
            E4 cds ->
      /\ 3' sj
3301 atgttctcct gcatctgtat ctagcactac acaagaagta tccattcctg aatctactac
3361 atacaccccc gcacagacct ccacccttgt gtcctcaagc accaaggaag acgcagtgca
3421 aacgccgcct aggaaacgag cacgaggagt ccaacagtcc ccttgcaacg ccttgtgtgt
3481 ggcccacatt ggacccgtgg acagtggaaa ccacaacctc atcactaaca atcacgacca
3541 gcaccaaaga cggaacaaca gtaacagttc agctacgcct aTAGtgcaat ttcaagGTga
                                                  <- E4 end
                                                          5' sj /\
3601 atccaattgt ttaaagtgtt ttagatatag gctaaatgac agacacagac atttatttga
3661 tttaatatca tcaacgtggc actgggcctc ctcaaaggca ccacataaac atgccattgt
3721 aactgtaaca tatgatagtg aggaacaaag gcaacagttt ttagatgttg taaaaatacc
3781 ccctaccatt agccacaaac tgggatttat gtcactgcac ctattgTAAt ttgtatatat
                                                     <- E2 end
3841 gtaaatgtgt aaatatatgg tattggtgTA Atacaactgt acatgtATGg aagtggtgcc
                        E5a orf start ->      E5a cds ->
3901 tgtacaaata gctgcaggaa caaccagcac attcatactg cctgttataa ttgcatttgt
3961 tgtatgtttt gttagcatca tacttattgt atggatatcT GAgtttattg tgtacacatc
                                     E5b orf start ->
4021 tgtgctagta ctaacactgc ttttatattt actattgtgg ctgctattaa caaccccctt
4081 gcaatttttc ctactaactc tacttgtgtg ttactgtccc gcattgtata tacactacta
4141 tattgttacc acacagcaAT GAtgctaaca tgtcaattta atgatggaga tacctggctg
               E5b cds -> <- E5a end
4201 ggtttgtggt tgttatgtgc ctttattgta gggatgttgg ggttattatt gatgcactat
4261 agagctgtac aaggggataa acacaccaaa tgtaagaagt gtaacaaaca caactgtaat
4321 gatgattatg taactatgca ttatactact gatggtgatt atatatatat gaatTAGagt
                                                                  <- E5b end
                                                     L2 orf start ->
4381 aaaccgtttt ttatatttgt aacAGtgtat gctttgtata ccATGgcaca tagtagggcc
                               /\ 3' sj     L2 cds ->
4441 cgacgacgca agcgtgcgtc agctacacag ctatatcaaa catgtaaact cactggaaca
4501 tgccccccag atgtaattcc taaggtggag cacaacacca ttgcagatca aatATTAAAA
                                                       signal ->
4561 tggggaagtt tgggggtgtt ttttggaggg ttgggtatag gcacgggttc cggcactggg
4621 ggtcgtactg gctatgttcc cttacaaact tctgcaaaac cttctattac tagtgggcct
4681 atggctcgtc ctcctgtggt ggtggagcct gtggcccctt cggatccatc tattgtgtct
4741 ttaattgaag aatcggcaat cattaacgca ggggcgcctg aaattgtgcc ccctgcacac
4801 ggtgggttta caattacatc ctctgaaaca actacccctg caatattgga tgtatcagtt
4861 actagtcaca ctactactag tatatttaga aatcctgtct ttacagaacc ttctgtaaca
4921 caaccccaac cacccgtgga ggctaatgga catatattaa tttctgcacc cactgtaacg
4981 tcacacccta tagaggaaat tcctttagat acttttgtgg tatcatctag tgatagcggt
5041 cctacatcca gtacccctgt tcctggtact gcacctcggc ctcgtgtggg cctatatagt
5101 cgtgcattgc accaggtgca ggttacagac cctgcatttc tttccactcc tcaacgctta
5161 attacatatg ataaccctgt atatgaaggg gaggatgtta gtgtacaatt tagtcatgat
5221 tctatacaca atgcacctga tgaggctttt atggacataa ttcgtttgca cagacctgcc
5281 attgcgtccc gacgtggcct tgtgcggtac agtcgcattg gacaacgggg gtctatgcac
5341 actcgcagcg gaaagcacat aggggcccgc attcattatt tttatgatat ttcacctatt
5401 gcacaggctg cagaagaaat agaaatgcac cctcttgtgg ctgcacagga tgatacattt
5461 gatatttatg ctgaatcttt tgaacctggc attaacccta cccaacaccc tgttacaaat
5521 atatcagata catatttaac ttccacacct aatacagtta cacaaccgtg gggtaacacc
5581 acagttccat tgtcacttcc taatgacctg tttttacaat ctggccctga tataactttt
5641 cctactgcac ctatgggaac acccctttagt cctgTAActc ctgctttacc tacaggccct
                                  L1 orf start ->
5701 gttttcatta caggttctgg attttatttg catcctgcat ggtattttgc acgtaaacgc
5761 cgtaaacgta ttcccttatt tttttcAGAT GtggcggccT AGcgacagca cagtatatgt
                               L1 cds ->           <- L2 end
                                     /\ 3' sj
5821 gcctcctcct aaccctgtat ccaaagttgt tgccacggat gcttatgtta ctcgcaccaa
5881 catattttat catgccagca gttctagact tcttgcagtg ggacatcctt atttttccat
5941 aaaacgggct aacaaaactg ttgtgccaaa ggtgtcagga tatcaataca gggtatttaa
6001 ggtggtgtta ccagatccta acaaatttgc attgcctgac tcgtctcttt tcgatcccac
6061 aacacaacgt ttagtatggg catgcacagg cctagaggtg ggcaggggac agccattagg
6121 tgtgggtgta agtggacatc ctttcctaaa taaatatgat gatgttgaaa attcagggag
```

```
 6181 tggtggtaac cctggacagg ataacagggt taatgtaggt atggattata aacaaacaca
 6241 attatgcatg gttggatgtg cccccccttt gggcgagcat tggggtaaag gtaaacagtg
 6301 tactaataca cctgtacagg ctggtgactg cccgccctta gaacttatta ccagtgttat
 6361 acaggatggc gatatggttg acacaggctt tggtgctatg aattttgctg atttgcagac
 6421 cAATAAAtca gatgttccta ttgacatatg tggcactaca tgtaaatatc cagattattt
signal ->
 6481 acaaatggct gcagacccat atggtgatag attatttttt tttctacgga aggaacaaat
 6541 gtttgccaga cattttttta acagggctgg cgaggtgggg gaacctgtgc ctgatacact
 6601 tataattaag ggtagtggaa atcgcacgtc tgtagggagt agtatatatg ttaacacccc
 6661 gagcggctct ttggtgtcct ctgaggcaca attgtttaat aagccatatt ggctacaaaa
 6721 agcccaggga cataacaatg gtatttgttg gggtaatcaa ctgtttgtta ctgtggtaga
 6781 taccacacgc agtaccaaca tgacattatg tgcatccgta actacatctt ccacatacac
 6841 caattctgat tataaagagt acatgcgtca tgtggaagag tatgatttac aatttatttt
 6901 tcaattatgt agcattacat tgtctgctga agtaatggcc tatattcaca caatgaatcc
 6961 ctctgttttg gaagactgga actttgggtt atcgcctccc ccaaatggta cattagaaga
 7021 tacctatagg tatgtgcagt cacaggccat tacctgtcaa aagcccactc ctgaaaagga
 7081 aaagccagat ccctataaga accttagttt ttgggaggtt aatttaaaag aaaagttttc
 7141 tagtgaattg gatcagtatc ctttgggacg caagtttttg ttacaaagtg gatatagggg
 7201 acggtcctct attcgtacag gtgttaagcg ccctgctgtt tccaaagcct ctgctgcccc
 7261 taaacgtaag cgcgccaaaa ctaaaaggTA ATATATGTGT ATATGTACTG TTATATATAT
                                     <- L1 end
                                       -> 24 bp tandem repeat  <-->
 7321 GTGTGTATGT ACTGTTATGt atatgtgtgt gtgtgttctg tgtgtaatGT aagttatttg
    24 bp tandem repeat<-                         5' sj /\
 7381 tgtaatgtgt atgtgtgttt atgtgcAATA AAcaattacc tcttgttaca ccctgtgact
                          signal ->
 7441 cagtggctgT TGCACGCGtt ttggtTTGCA CGCGccttac acacataagt aatatacatg
       repeat ->         repeat ->
 7501 cacaatatat atatttttgt ttaaaatact atacttttat atttgcaacc gttttcggtt
 7561 gcccttagca tacactttcc accaatttgt tacaacgtgt ttcctctTAA tcctatatat
                                               E8 orf start ->
 7621 tttgtgccag gtacacattg ccctgccaag ttgcttgcca agtgcatcat atcctgccaa
 7681 ccacacacct ggcgccaggg tgcggtattg ccttactcat aaacctgtct ttgtgttata
 7741 cttttATGca ctgtagccaa ctcttaaaag cattttggc ttgtagcagc acatttttt
   E8 cds ->
 7801 gctcttactg tttggtatac aataacataa aaatgagtaa cctaaggtca cacacctgcg
 7861 accggtttcg gttatccaca ccctacatat ttccttcTTA TA
                                            -> CAAT-box start
```

```
LOCUS       HPV11        7931 bp ds-DNA   circular  VRL       30-SEP-1988
DEFINITION  Human papillomavirus type 11 (HPV-11), complete genome.
ACCESSION   M14119
SOURCE      Human laryngeal papillomavirus type 11 DNA recovered from a
            laryngeal papilloma.
REFERENCE   1  (bases 1 to 7931)
  AUTHORS   Dartmann,K., Schwarz,E., Gissmann,L. and Zur Hausen,H.
  TITLE     The nucleotide sequence and genome organization of human papilloma
            virus type 11
  JOURNAL   Virology 151, 124-130 (1986)
COMMENT     HPV-11 and HPV-6 are responsible for the large majority of
            exophytic condylomas in the genital tract.  Even though these
            lesions are frequently present in the genital tract, they are
            virtually absent in higher grade neoplasias and in cervical
            cancers.  HPV-11  also infects other mucosal types; the respiratory
            tract, oral cavity and conjunctiva.  It has been recovered from
            approximately 50% of respiratory tract lesions and 50% of all
            childhood conjunctival papillomas.  Respiratory papillomatosis is
            a rare disease that can be life-threatening because of its
            recurrent nature and the possibility of obstruction of the airways
            and respiratory distress.  The most frequent sites of infection are
            the vocal cords in the larynx, but papillomas may also be present
            in the trachea, lungs, nose and oral cavity.  These respiratory
            papillomas progress to malignancy rarely, as they account for less
            than 0.1% of all respiratory cancers.

            The complete genome of HPV11 was cloned from the laryngeal
            papilloma of a 15-year-old patient.  The sense strand, containing
            all ORF's of significant length, is shown, with the first position
            of the circular genome defined by alignment to HPV6b.  Designation
            of early and late ORF's is based on homology with HPV6b, to which
            it is 82% similar at the nucleotide level.  A TATA box is shown
            36 bp upstream from the beginning of the E6 cds and a corresponding
            CAAT box is found 57 bp upstream from the TATA box.
            Polyadenylation signals can be found mostly following the late
            coding regions.  A 12-bp-long palindromic sequence responsible for
            binding of the E2 protein can be found repeated four times in the
            non-coding LCR region between L1 and E6.  Also housed in the LCR
            are several short direct repeats that are up to 16 bp in length.
            The authors also note that the first 200 nucleotides of the LCR has
            an unusually high T content (53%) and an unusually low C content
            (2%).

BASE COUNT     2406 a   1519 c   1736 g   2270 t
ORIGIN      4557 bp upstream of HindIII site; 101 bp upstream from beginning
            of E6 cds
        1 cttaataaCA ATCTTAGTTT AAAAaagagg agggACCGAA AACGGTtcaA CCGAAAACGG
           signal ->        -> signal          -> E2 bind      -> E2 bind
              E6 orf start ->
       61 TtataTATAA Accagcccaa aaaattagca gacgaggcat tATGgaaagt aaagatgcct
        signal ->   <-                          E6 cds ->
      121 ccacgtctgc aacatctata gaccagttgt gcaagacgtt taatctttct ttgcacactc
      181 tgcaaattca gtgcgtgttt tgcaggaatg cactgaccac cgcagagata tatgcatatg
      241 cctataagaa cctaaaggtt gtgtggcgag acaactttcc ctttgcagcg tgtgcctgtt
      301 gcttagaact gcaagggaaa attaaccaat atagacactt taattatgct gcatatgcac
      361 ctacagtaga agaagaaacc aatgaagata ttttaaaagt gttaattcgt tgttacctgt
      421 gtcacaagcc gttgtgtgaa atagaaaaac taaagcacat attgggaaag gcacgcttca
      481 taaaactaaa TAAccagtgg aagggtcgtt gcttacactg ctggacaacA TGcatggaag
           E7 orf start ->                              E7 cds ->
      541 acttgttacc cTAAaggata tagtactaga cctgcagcct cctgaccctg tagggttaca
                      <- E6 end
      601 ttgctatgag caattagaag acagctcaga agatgaggtg gacaaggtgg acaaacaaga
      661 cgcacaacct ttaacacaac attaccaaat actgacctgt tgctgtggat gTGAcagcaa
                                                     E1 orf start ->
      721 cgtccgactg gttgtggagt gcacagacgg agacatcaga caactacaag acctttgct
```

```
 781 gggcacacta aatattgtgt gtcccatctg cgcaccaaaa ccaTAAcaag gATGgcggac
                                                    E1 cds ->
                                                   <- E7 end
 841 gattcaggta cagaaaatga ggggtcgggg tgtacaggat ggtttatggt agaagccata
 901 gtagagcaca ctacaggtac acaaatatca gaagatgagg aagaggaggt ggaggacagt
 961 gggtatgaca tggtggactt tattgatgac aggcatatta cacaaaattc tgtggaagca
1021 caggcattgt ttaataggca ggaggcggat gctcattatg cgactgtgca ggacctaaaa
1081 cgaaagtatt taggcagtcc atatgtaagt cctataagca atgtagctaa tgcagtagaa
1141 agtgagataa gtccacggtt agacgccatt aaacttacaa cacagccaaa aaaggtaaag
1201 cgacggctgt ttgaaacacg ggaattaacg gacagtggat atggctattc tgaagtggaa
1261 gctgcaacgc aggtagagaa acatggcgac ccggaaaatg ggggagatgg tcaggaaagg
1321 gacacaggga gggacataga gggtgagggg gtggaacata gagaggcgga agcagtagac
1381 gacagcaccc gagagcatgc agacacatca ggaatattag aattactaaa atgtaaggat
1441 atacgatcta cattacatgg taagtttaaa gactgctttg ggctgtcatt tgttgattta
1501 attaggccat ttaaaagtga tagaaccaca tgtgccgatt gggtggttgc aggatttggt
1561 atacatcata gcatagcaga tgcatttcaa aagttaattg agccattaag tttatatgca
1621 catatacaat ggcttacaaa tgcatgggga atggtactat tagtattaat aaggtttaaa
1681 gtaaataaga gcagatgtac cgtggcacgt acattaggta cgttattaaa tatacctgaa
1741 aatcacatgt taattgagcc tcctaaaata caaagtggcg tacgagccct gtattggttt
1801 aggacaggca tttcaaatgc aagtacagtt ataggggagg cgccggaatg gataacgcgc
1861 cagaccgtta ttgaacatag tttggctgac agtcaattta aattaactga aatggtgcag
1921 tgggcatatg ataatgatat ttgtgaagaa agtgagatag catttgaata tgcacagcgt
1981 ggagactttg actccaatgc aagggccttt ttaaatagta atatgcaggc taaatatgta
2041 aaagattgtg caattatgtg cagacattat aaacatgcag aaatgaaaaa gatgtctatt
2101 aaacaatgga ttaagtatag gggtactaaa gttgacagtg taggtaactg gaagccaatt
2161 gtgcagtttc taagacatca aaacatagaa tttattccat ttttaagcaa actaaaatta
2221 tggctgcacg gaacgcccaa aaaaaattgt atagccattg tagggccacc tgacactggg
2281 aagtcgtgct tttgcatgag tttaattaag tttttggggg gaacagttat tagttatgtt
2341 aattcctgca gccatttctg gctacagcca ctaacggatg caaaagtggc attattggat
2401 gatgccacac aaccatgttg gacatatatg gatacatata tgagaaacct attagatggt
2461 aatcctatga gcatagatag aaaacataga gcattaacat taattaagtg tccaccgcta
2521 ctggttacat caaatataga cattagcaaa gaggagaaat acaaatattt acatagtaga
2581 gttaccacat ttacatttcc aaatccattc ccctttgaca gaaatgggaa tgcagtatat
2641 gaactatcag atgcaaactg gaaatgtttc tttgaaagac tgtcgtccag ccTAGacatt
                                                      E2 orf start ->
2701 gaggattcag aggacgagga agATGgaagc aatagccaag cgtttagatg cgtgccagga
                     E2 cds ->
2761 tcagttgtta gaactttaTG Aagaaaacag tattgatata cacaaacaca ttatgcattg
                           <- E1 end
2821 gaaatgcata cgattggaaa gtgtattact acacaaagca aaacaaatgg gcctgagcca
2881 catcgggtta caagtagtac caccattaac tgtgtcagag actaaaggac ataatgctat
2941 tgaaatgcaa atgcatttag aatccttagc aaaaactcag tatggtgtgg aaccttggac
3001 attacaggac accagttatg aaatgtggct aacaccaccc aaacggtgct ttaaaaaaca
3061 gggaaatact gtggaggtaa aatttgatgg ctgtgaagac aatgtaatgg agtatgtggt
3121 atggacacat atatacctgc aggacaacga ctcatgggta aaagtaacta gttccgtaga
3181 tgccaagggc atatattata catgtggaca atttaaaaca tattatgTAA attttAATAA
                                                 E4 orf start ->   -> signal
3241 Agaggcacaa aagtATGgta gtaccaatca ttgggaagta tgttatggca gcacagttat
            E4 cds ->
3301 atgttctcct gcatctgtat ctagcactgt acgagaagta tccattgctg aacctactac
3361 atacaccccc gcacagacca ccgcccctac agtgtccgcc tgcaccacgg aagacggcgt
3421 gtcggcgccg cctaggaagc gagcacgtgg accgtccact aacaacaccc tgtgtgtggc
3481 caacatcaga tccgtggaca gtacaatcaa caacatcgtc actgacaatt acaacaagca
3541 ccaaagaagg aacaactgtc acagtgcagc tacgcctaTA Gtgcaactgc aaggtgattc
                                                  <- E4 end
3601 caattgttta aaatgtttta gatatagact gaatgacaaa tataaacatt tgtttgaatt
3661 agcatcttca acgtggcatt gggcctcacc tgaggcacca cataaaaatg caattgtaac
3721 attaacatat agcagtgagg aacaacgtca gcaattttta aacagtgtaa aaataccacc
3781 caccattagg cataaggtgg ggtttatgtc attacattta ttgTAAccat tacacctgta
                                                      <- E2 end
3841 tatatgtata tgtgtacaTA Acatacgtgt ATGgaggtag tgcctgtaca aattgctgca
              E5A orf start ->
                                E5A cds ->
3901 gcaacaacta caacattgat attgcctgtt gttattgcat ttgcagtatg tattcttagt
```

```
3961 attgtactta taatattaat atcTGAtttt gtagtatata catctgtgct ggtactaaca
                  E5B orf start ->
4021 cttcttttat atttgctttt gtggctttta ttaacaaccc ctttgcaatt ctttttacta
4081 acactgtgtg tgtgctattt tcctgccttt tatatacaca tatacattgt gcaaacgcaa
4141 caaTAATGgt gatgttaacc tgtcacttaa atgatggtga tacatggttg tttctgtggt
  E5B cds ->
        <- E5A end
4201 tgtttactgc atttgttgta gctgtacttg gattgttgtt actacattac agggctgtac
4261 atggtactga aaaaactaaa tgtgctaagt gtaaatcaaa ccgcaatact actgtggatt
4321 atgtgtatat gtcacatggt gataatggag attatgtgta catgaacTAG AGTAAAcctt
                                                        <- E5B end
                                                    signal ->
4381 ttttatacag tgtgtggtgt acgtTAGtta tatataATGa aacctagggc acgcagacgt
                     L2 orf start -> L2 cds ->
4441 aaacgtgcgt cagccacaca actatatcaa acatgcaagg ccactggtac atgtccccca
4501 gatgtaattc ctaaagttga acatactact attgcagatc aaatATTAAA Atggggaagc
                                              signal ->
4561 ttaggggttt tttttggtgg gttaggtatt ggtacagggg ctggtagtgg cggtcgtgca
4621 gggtatatac ccttgggaag ctctcccaag cctgctatta ctgggggccc agcagcacgt
4681 ccgccagtgc ttgtggagcc tgttgcccct tccgatccct ccattgtgtc cttaattgag
4741 gagtctgcta ttattaatgc tggtgcacct gaggtggtac cccctacaca gggtggcttt
4801 actataacat catctgaatc gactacacct gctattttag atgtgtctgt taccaatcac
4861 actaccacta gtgtgtttca aaatcccctg tttacagaac cgtctgtaat acagccccaa
4921 ccacctgtgg aggccagtgg tcacatactt atatctgccc caacaataac atcccaacat
4981 gtagaagaca ttccactaga cacttttgtt gtatcctcta gtgatagtgg acctacatcc
5041 agtactcctc ttcctcgtgc ttttcctcgg cctcgggtgg gtttgtatag tcgtgcctta
5101 cagcaggtac aggttacgga ccccgcgttt ttgtccacgc cacagcgatt ggtaacttat
5161 gacaaccctg tctatgaagg agaagatgta agtttacaat ttacccatga gtctatccac
5221 aatgcacctg atgaagcatt tatggatatt attagactac atagaccagc tataacgtcc
5281 agacggggtc ttgtgcgttt tagtcgcatt gggcaacggg ggtccatgta cacacgcagt
5341 ggacaacata taggtgcccg catacattat tttcaggaca tttcaccagt tacacaagct
5401 gcagaggaaa tagaactgca ccctctagtg gctgcagaaa atgacacgtt tgatatttat
5461 gctgaaccat ttgaccctat ccctgaccct gtccaacatt ctgttacaca gtcttatctt
5521 acctccacac ctaataccct ttcacaatcg tggggtaata ccacagtccc attgtcaatc
5581 cctagtgact ggtttgtgca gtctgggcct gacataactt ttcctactgc atctatggga
5641 acacccttta gtcctgTAAc tcctgcttta cctacaggcc ctgtttttat tacaggttct
           L1 orf start ->
5701 gacttctatt tgcatcctac atggtacttt gcacgcagac gccgtaaacg tattccctta
5761 ttttttacag ATGtggcggc cTAGcgacag cacagtatat gtgcctcctc ccaaccctgt
         L1 cds ->           <- L2 end
5821 atccaaggtt gttgccacgg atgcgtatgt taaacgcacc aacatatttt atcatgccag
5881 cagttctaga ctccttgctg tgggacatcc atattactct atcaaaaaag ttaacaaaac
5941 agttgtacca aaggtgtctg gatatcaata tagagtgttt aaggtagtgt tgccagatcc
6001 taacaagttt gcattacctg attcatccct gtttgacccc actacacagc gtttagtatg
6061 ggcgtgcaca gggttggagg taggcagggg tcaaccttta ggcgttggtg ttagtgggca
6121 tccattgcta aacaaatatg atgatgtaga aaatagtggt gggtatggtg gtaatcctgg
6181 tcaggataat agggttaatg taggtatgga ttataaacaa acccagctat gtatggtggg
6241 ctgtgctcca ccgttaggtg aacattgggg taagggtaca caatgttcaa atacctctgt
6301 acaaaatggt gactgccccc cgttggaact tattaccagt gttatacagg atggggacat
6361 ggttgataca ggctttggtg ctatgaattt tgcagactta caaaccAATA AAtcggatgt
                                                 signal ->   <-
6421 tccccttgat atttgtggaa ctgtctgcaa atatcctgat tatttgcaaa tggctgcaga
6481 cccttatggt gataggttgt ttttttattt gcgaaaggaa caaatgtttg ctagacactt
6541 ttttaatagg gccggtactg tgggggaacc tgtgcctgat gacctgttgg taaaaggggg
6601 taataacaga tcatctgtag ctagtagtat ttatgtacat acacctagtg gctcattggt
6661 gtcttcagag gctcaattat ttAATAAAcc atattggctt caaaaggctc agggacataa
                    signal ->  <-
6721 caatggtatt tgctggggaa accacttgtt tgttactgtg gtagatacca cacgcagtac
6781 aaatatgaca ctatgtgcat ctgtgtctaa atctgctaca tacactaatt cagattataa
6841 ggaatacatg cgccatgtgg aggagtttga tttacagttt atttttcaat tgtgtagcat
6901 tacattatct gcagaagtca tggcctatat acacacaatg aatccttctg ttttggagga
6961 ctggaacttt ggtttatcgc ctccaccaaa tggtacactg gaggatactt atagatatgt
7021 acagtcacag gccattacct gtcagaaacc cacacctgaa aaagaaaaac aggatcccta
7081 taaggatatg agtttttggg aggttaactt aaaagaaaag ttttcaagtg aattagatca
```

```
7141 gtttcccctt ggacgtaagt ttttattgca aagtggatat cgaggacgga cgtctgctcg
7201 tacaggtata aagcgcccag ctgtgtctaa gccctctaca gcccccaaac gaaaacgtac
7261 caaaaccaaa aagTAAtata tgtgtgtcag tgtgttgtgt tatttatatg ttgttgtagT
                  <- L1 end                    repeat region start ->
7321 GTGTATATGT TTcttgtaTT GTGTATATGT GTATATGTTT GTGTATATGT GTATGTTATG
7381 TATGTTatgt TGTTATGTAT GTTtgtgtgt ttagtgtgtg tatatatttg tggaatgtgt
         repeat region end <-
7441 atgtatgttt ttgtgcAATA AAcaattatt atgtgtgtcc tgttacaccc agtgactaag
              signal ->   <-
7501 ttgtgttttg cacgcgccgt ttgtgttgcc ttcatattat attatatata tttgtaatat
7561 acctatacta tgttaccccc ccccacttgc aACCGTTTTC GGTtgccctt acatacactt
                                       -> E2 bind
7621 acctcaaatt tgttataacg tgttttgtac taatcccata tgttgtgtgc caaggtacat
7681 attgccctgc caagtatctt gccaacaaca cacctggcca gggcgcggta ttgcatgact
7741 aatgtacAAT AAAcctgtcg gtttgtacaa tgttgtggat tgcagccaaa ggttaaaagc
     signal ->
7801 atttttggct tctagctgaa cattttgta cccttagtat attatgcaca atacccacaa
7861 aatgagtaac ctaaggtcac acacctgcaA CCGGTTTCGG Ttacccacac cctacatatt
                                    -> E2 bind
7921 tccttcttat a
```

```
LOCUS       HPV13        7880 bp ds-DNA             VRL       01-DEC-1992
DEFINITION  Human papilloma virus type 13 (HPV-13), complete genome.
ACCESSION   X62843
KEYWORDS    genome.
SOURCE      Human papillomavirus type 13 DNA recovered from FEH lesions from a
            13 year old Turkish girl.
REFERENCE   1  (bases 1 to 7880)
  AUTHORS   Van Ranst,M., Fuse,A., Fiten,P., Beuken,E., Pfister,H.,
            Burk,R.D. and Opdenakker,G.
  TITLE     Human papillomavirus type 13 and pygmy chimpanzee papillomavirus
            type 1: Comparison of the genome organizations
  JOURNAL   Virology 190, 587-596 (1992)
COMMENT     HPV13 belongs to a subgroup of orogenital papillomaviruses
            (including also HPV6, HPV11, HPV43, HPV44) associated as a group
            with benign orogenital lesions (FEH, condyloma acuminata and
            low-grade cervical neoplasia), and only rarely with cervical
            cancer.  In addition to its similarity to other HPVs, HPV-13 has
            been shown to be 85% similar to a pigmy chimpanzee
            papillomavirus (PCPV-1) cloned from a colony of pigmy chimpanzees
            afflicted by a FEH-like disease.  Van Ranst et al. contend that
            this close relationship indicates the possibility of cross-species
            transmission.

            HPV13 itself is strongly associated with oral focal epithelial
            hyperplasia (FEH).  FEH is characterized by multiple and discrete
            nodular elevations of the oral mucosa.  Its prevalence is high in
            Indians in Central and South America and in Eskimos in Greenland
            and Alaska, but is low in Caucasians who live in the same area.
            Recently, HPV-13 has been detected by PCR in a case of low-grade
            cervical dysplasia and by in situ hybridization in Bowenoid
            papulosis in an HIV-positive male.  In both of these instances,
            HPV16 was also present in the lesion.  Van Ranst et al. contend
            that the discovery of HPV13 in these recent cases along with the
            relative rarity of FEH in the Caucasian population to which the
            subjects belonged may point to the existence of a hidden reservoir
            of virus in this population.

            The complete genome of HPV-13 was cloned from the FEH lesions of a
            13-year-old Turkish girl.  The sense strand, containing all ORF's
            of significant length, is shown, with the first position of the
            circular genome defined by homology with HPV6 and HPV11.  The E6
            ORF of HPV-13 contains four copies of the cysteine doublet motif.
            This protein may be involved with the binding of p53.  Also this
            ORF, as is the case with types 6, 11, and 44, does not contain the
            splice acceptor/donor pair which seems to be critical for E7
            expression.  The E7 ORF of HPV-13 contains a degenerate form of the
            cell division motif which may mediate binding of the tumor
            suppressor protein pRB-105.  In the beginning of the E2 ORF, HPV-13
            contains a putative leucine zipper motif.  This motif has been found
            to be degenerate in the high-risk papillomaviruses.  It has been
            speculated that this motif could mediate the binding of the E2
            activator to the URR, which blocks transcription of E6.

            The upstream regulatory region (URR) contains several promoter
            elements and transcription factors: TATA box at nt 68, CAAT signal
            at nt 9, 5 NF-1 binding sites at 7587, 7607, 7717, 7741, and 7761,
            2 AP-1 binding sites at 7404 and 7810, Oct-1 binding site at 7324,
            and 4 copies of the E2 binding region.  Unlike many other genital
            papillomaviruses, HPV-13 does not house a glucocorticoid response
            element.  Thus, van Ranst et al. believe it may not be influenced
            by steroid hormones.

BASE COUNT     2452 a   1480 c   1628 g   2320 t
ORIGIN      degenerate HpaI site; 103 bp upstream from beginning of E6 cds
        1 gtttctaaCA ATCTtaagtt taaaaaatag gtgggACCGA AAACGGTttT AACCGAAAAC
           signal ->                                E6 orf start ->
```

HPV13

```
                                    E2 bind ->        E2 bind ->
    61 GGTgataTAT AAAccagccc aaaaattgag caagcggggc ataATGgaaa gtgcaaatgc
       signal ->                              E6 cds ->
   121 ctccacgcct gcaaaaacta tagaccagtt gtgcaaggag tgcaaccttt ctatgcacag
   181 cttgcaaatt ctatgcgtgt tctgcaggaa aaccctgtcc acggcagagg tttatgcatt
   241 tcagtataag agtttatata tagtgtggcg aggacagttt ccatttgcgg cttgtgcatg
   301 ctgcttagaa atacaaggaa agattaacca gtttaggcat tttgacttcg cgggatttgc
   361 tgtaacagtt gaagaagaca caaagcagtc aattttggat gtgctaattc gctgctattt
   421 atgccacaaa ccattgtgtg aagtggagaa actaagacat attttgcaga aggcacgatt
   481 tatTAAatta aacagcagtt ggaaaggccg ctgttttcat tgctggtcat cATGcatgga
E7 orf start ->                                      E7 cds ->
   541 aaatatccta cctTAAaaga cattgtttta gagctgactc ctgaccctgt aggtctacat
                     <- E6 end
   601 tgcaatgagc aattagacag ctcagaagac gaggtggacg aacaagccac gcaagccacg
   661 caagccacgc aacatagcac actattacaa tgctaccaaa tactaacgtc ctgtagtaaa
   721 tgttgtagca acgtccggct ggtggtggag tgtacaggac cTGAcattca cgacctacac
                                       E1 orf start ->
   781 gacctactgc tgggcacgct gaatatagtg tgccctttgt gtgcaccaaa aagcTAAcca
                                                                  <- E7 end
   841 cgATGgcaga ggatacaggt actaataatg aggggacggg atgctcagga tggtttttag
  E1 cds ->
   901 tagaggctgt agtagaacga acaactgggc aacaaatatc agatgatgag gatgaaacag
   961 tggaagatag tgggttggat atggtggatt tcatagatga cagacctatt acacacaatt
  1021 ccgtggaagc acaggcattg ttaaacgagc aggaggcgga tgctcattat gcggctgtgc
  1081 aggacctaaa acgaaagtat ttaggcagtc catatgttag tcccctagga catgttgaac
  1141 agtcagtgga ctgtgatata agtcctcgat tggacgctat aaaattaagt agaaattcta
  1201 aaaaagtaaa gcgacggctg tttcaatcaa gggaaataac ggacagtgga tatggctatt
  1261 ctgaagtgga agctgaaacg caggtagaga gaaatggcga accggaaaat gattgtgggg
  1321 gtggtggaca cggaagggac aaagaggggg agggacaggt gcacacggaa gtgcacacag
  1381 gcagccagat agaagagcac acagggacca cgcgggtgtt agaactcctt aaatgtaagg
  1441 atgtaagggc tacattgtat ggtaagttta aagactgtta tgggttatca tttacagatt
  1501 taattagacc atttaaaagt gataaaacaa catgtgggga ctgggtggtt gcagcatttg
  1561 gtatacatca tagtgtatca gaggcatttg aaaagttaat gcagccatta acaacatata
  1621 tgcatataca atggcttaca aatgcatggg ggatggtatt gttagtatta ataagattta
  1681 aagtaaataa aagtagatgc acagtggcgc gaacactggc aacctttctt aatattcctg
  1741 aggaccacat gttaattgaa cctcccaaaa tacaaagcag tgtggcagca ttatactggt
  1801 ttagaacagg tatttctaat gctagtatag taactggtga aacaccagaa tggataaaaa
  1861 ggcaaacaat tgtagagcat ggacttgcag ataatcaatt taaattaact gaaatggtgc
  1921 agtgggcata tgataatgat ttttgtgatg aaagcgaaat agcatttgaa tatgcacaac
  1981 gaggagattt tgattcaaat gccagggcat ttttaaatag taattgtcag gcaaaatatg
  2041 taaaagattg tgcaacaatg tgcaagcatt ataaaaatgc agaaatgaaa aaaatgtcta
  2101 tgaaacaatg gataacatat agaagtaaaa aaatagagga agcaggaaat tggaaaccaa
  2161 tagtacaatt tttaaggcat caaaatatag aatttattcc atttttaagt aaattaaaat
  2221 tgtggcttca tggcacgcca aagaaaaact gtattgcaat agtggggcca ccagatacag
  2281 gcaaatcatg tttttgcatg agcttaatta agtttttagg gggcacagta attagttatg
  2341 taaattcaag tagccatttt tggctgcagc cattatgtaa tgcaaaggta gctttgctag
  2401 atgatgcaac gcagtcatgc tgggtatata tggacacata catgagaaat ttattagatg
  2461 gcaatccaat gagcattgat agaaaacata agtctttagc attaataaaa tgtccgccat
  2521 tattagtaac atctaatgta gacattacca aagatgacaa atataaatat ttgtatagta
  2581 gagtaacaac acttacattt ccaaatccat tccctttga cagaaatggg aatgcagtat
  2641 atgagttgtc tgatgcaaac tggaaatgtt tttttacaag attgtcagca agccTAGata
                                                       E2 orf start ->
  2701 tacaggactc tgaggacgag gacgATGgag acaatagcca agcatttaga tgcgtgccag
                          E2 cds ->
  2761 gaacagttgt tagaactgta TGAagaaaat agtaatgaac ttaaaaaaca tatacaacat
                            <- E1 end
  2821 tggaaatgct taaggtacga aagtgtactc ttacacaaag cacgccaaat gggcctaagc
  2881 cacattggat tacaagtggt gccaccattg acagtatcac aagctaaggg acatgaggca
  2941 attgaaatgc aaatgacttt agagacatta ctagagtctg agtttggtat ggaaccatgg
  3001 actttacaag atacaagtcg tgaaatgtgg ctaacacccc caaaacgctg ttttaagaaa
  3061 cagggacaaa ctgtggaagt aaaatatgac tgtaatacag acaatagaat ggattatgtg
  3121 tcgtggacat acatatatgt gtttgacaca gataaatgga caaaggtgaa aggaatggta
  3181 gattataaag ggttgtacta catacatgga aatttgaaaa catattattT AGagtttgaa
                                                  E4 orf start ->
```

```
   3241 aaggaggcta aaaaatATGg ggaaacgtta caatgggaag tatgtattgg cagcacagtc
                 E4 cds ->
   3301 atatgttctc ctgcatctgt atctagtact gtacaagaag tatccattgc tgggcctgct
   3361 tcatactcca ccaccacctc cacacaggcc tccaccgcag tgtcctgcag cgcctcggaa
   3421 gaatgtgtgc aagcgccgcc ttgtaaacga caacgaggac cttcacgtcc cattggaaac
   3481 ccccagaaca cacaaagcat tgtgtgtgtc acagactacg acaccctgga cagtgcaaac
   3541 aacaacatca acgttaacca ttacaacaat aacaaaggac gggacaacag ttactgtgca
   3601 gctacaccta TAGttcaatt acaaggtgac tctaattgtc taaagtgttt tcgatataga
                  <- E4 end
   3661 ttacatgaaa aatataaaga tttatttttg ttagcatcat ctacatggca ttggaccgcc
   3721 cctaataatt cacaaaaaca tgcactggta accttaacct atgtaaatga acaacaaaga
   3781 caagactttt taaaaactgt aaaaatacct ccaaccataa cacataaact aggttttatg
   3841 tcattgcaat tgttaTAAca gcatatattg tatgtaaata tttgttgtgt gtgtgtatat
                        <- E2 end
   3901 attgTAAATG gaatttatac ctgtggatgt tagtacacag gcaaccagca agtcattact
E5 orf and cds ->
   3961 gccacttgta attgcactta cagtgtgtgt agttagcatt ataacaatat tgtgcatatc
   4021 agagttcttg gtgtacacaa acgttttagt actaacatta attttatatg tacttttgtg
   4081 gcttttacta acaactccct tgcaattcta tttactaacc ctgtctcttt gctttcttcc
   4141 tgcgttgtgt gtacaccaat atattttaca aacacaagaa TAActataca caatgttaac
                                                    <- E5 end
   4201 ctgtactttt gatgatggtg acacatggtt gctattatgg ttaattttat catttattgt
   4261 agccattcta gggttactgt tgctgtatat aagaactgga catatgcatt gccagtgctg
   4321 gagtaaaTAA gtggttttat atttgtgtg tattcattta attATGgcac atagtagggc
      L2 orf start ->                          L2 cds ->
   4381 tcgcagacgc aaacgcgctt cagctacaca actatatcaa acttgtaagg cttctggaac
   4441 atgtcctcct gatgttatac caaaggttga acaaaacact cttgcagata aaatATTAAA
                                                          signal ->
   4501 gtggggcagt ttaggagtat tttttggggg gcttggcatt ggcacaggct ctggtactgg
   4561 cggtaggact ggctatgtac cagtaggatc caccccacgc cctgccatat caactgggcc
   4621 tactgcacgt cctcctattg ttgttgatac tgttgggcct acagaccctt ctattgtatc
   4681 tttggtagag gaatcagcta ttattaattc tggagtacct gacccttgc ctcccgttca
   4741 tgggggtttt gaaatcacca catctcaatc agccactcca gcaatattgg atgtgtctgt
   4801 tacaacacaa aacactacgt ccacaagtat atttagaaat cctgtttttt cagaaccttc
   4861 tattacacaa tctcaacctt ctattgaaag tggtgcacac gtgtttatat cgccatctac
   4921 tatttcccct cattctacag aagacattcc tttagataca tttattgtat cttcctcaga
   4981 tagtaatcct gcatcaagca cccctgttcc agcaactgtt gcacgtccac gtctaggcct
   5041 ttacagtagg gccttacatc aagtacaggt tactgatcct gccttttttat cgtcgcccca
   5101 acgccttata acctttgata accctacata tgaaggtgaa gatataagtt tgcagtttgc
   5161 acacaatacc attcatgaac cccctgatga ggcatttatg gatattataa gactacatag
   5221 gccagccata acatcacggc gtggtcttgt taggtttagt agaattggtc agaggggtc
   5281 tatgtatact cgaagcggca agcatatagg tggaagggtc catttcttta aggatatttc
   5341 tcctatatct gcagctgcag aagaaataga attacacccc cttgtggctg ctgcacagga
   5401 tcacagtggt ttgtttgata tttatgcaga acctgaccct gaccctgtgg ctgtaaacac
   5461 ctctgggtca ttgtcttctg cctccacacc atttgcacaa tcttctttgt cttccgcccc
   5521 atggggtaat actactgttc ctctttcact accaggtgat atatttatac agcctggtcc
   5581 tgacataaca ttcccaactg cacctacagT AAcgccttat aatcctgtta cgcctgcttt
                             L1 orf start ->
   5641 acctacaggt cctgttttta ttactgcttc tggattttat ttatatccta catggtattt
   5701 tacacgcaaa cgccgtaaac gtgtttcctt gtttttttaca gATGtggcgg ccTAGtgaca
                                            L1 cds ->           <- L2 end
   5761 acaaactata tgtgcctcct cccgcccctg tatcaaaagt aattactacg gatgcctatg
   5821 ttacacgtac caacatattt tatcatgcta gcagttctag actacttgca gtgggaaatc
   5881 cttattttcc tattaagaaa caaaacaaaa ctgttgtccc taaggtatct ggttatcagt
   5941 ttagggtatt taaagttgta ttacctgacc ctaataaatt tgccctgcct gacacatcta
   6001 tatttgactc aactagtcaa cgcttagtgt gggcctgtac aggtttagag gttggtaggg
   6061 gtcaaccctt aggtgttggt attagtggtc atccattatt aaataaatat gatgatgtgg
   6121 aaaattctgc aagttatgct gccaatcctg gtcaggataa tagggttaat gtggccatgg
   6181 actataaaca aacacagtta tgtttagtgg gctgtgcacc tcctttaggt gaacattggg
   6241 gacagggcaa gcaatgtact ggtgtaaatg tacaacctgg agattgccct cctttagaat
   6301 taattagtag tgtaattcag gatggtgaca tggtggatac aggatttgga gccatgaatt
   6361 ttgcggaatt gcaatctAAT AAAtctgatg tgccactaga catatgcacg tccacatgca
                 signal ->
   6421 aatatcctga ctatttacaa atggctgcgg atccttatgg agacagatta ttttttatc
```

```
6481 tgcgaaagga acaaatgttt gcaaggcatt tctttaacag ggcaggctct gttggtgaac
6541 aaatcccagc agaattatat gttaagggta gtaatacact ttctaatagt atttactata
6601 atactcccag tggctctctt gtgtcttctg aggcccagtt gtttAATAAA ccttattggt
                                              signal ->
6661 tacaaaaggc ccagggacac aataatggta tatgttgggg caatcacttg tttgttactg
6721 tagttgatac tacacgcagt actaacatga ctgtgtgtgc agccactaca tcatctcttt
6781 cagacacata taaggccaca gaatataaac agtacatgcg acatgtagaa gaatttgatt
6841 tacaatttat ttttcaattg tgcactatta aattaactgc agaggttatg tcatatattc
6901 atactatgaa tcctacaatt ctagaagact ggaactttgg gctatctccc cctcctaatg
6961 gaacattaga agacacatat agatatgtac aatctcaggc cataacgtgt caaaagccta
7021 cacctgataa agaaaaacag gatccgtatg cgggtcttag tttttgggag gttaatctta
7081 aggaaaagtt ttctagtgaa ctagatcagt atccccttgg cagaaagttt ttattacaaa
7141 caggcgttca gtctaggtcc cctattcgtg taggtaggaa acgtgctgca tctacatcta
7201 ctgccacacc tactacacgt aaaaaagcta aaaggaaaTA Atagtttgtt tatgattgtg
                                                <- L1 end
7261 tatgtatgtc acgtttgttt gtactgtatg tatgttgtgt actgtatgtg taatgttgta
7321 tgtATGTGCA Tgttacttat taaagaatgt gtgtgtgtgt ttgtatgcAA TAAAtctaat
       -> potential Oct-1 bind                   signal ->
7381 ctgtggtgtc ctgttccacc ctaTGAGTAA gtggtatgtt gtgtctcgtg tggtgttttg
          putative AP-1 bind ->
7441 tatactatac tataacatta gtgcaaccat tttgtaactt ttcttacatt ttacgtctcc
7501 atattaagtg caACCGATTT CGGTtgctat tgtttctgcg accgatttgt tgcagcacgc
         E2 bind ->
7561 tgtttatata atcttaccta ccgcctGCCA AAattatcca ccgcttGCCA AAatcaccca
                            NF1 ->               NF-1 ->
7621 cacacctggc gttgctaggg cgcggttata tatatttact aaatcttact aatctttcta
7681 tcactcattt tacctttata acaatacttt tgctttTCAA GTacattttt gtacttacta
                                      NF-1 ->
7741 GCCAATgcct gaaaggtttt TTGGCTacca gcactacatt tttgtacagt taatgttaca
NF-1 ->               NF-1 ->
7801 tgtataaaaT GAGTAAccta aggtcacaca cctgcaaACC GGTATCGGTt aaaacacacc
             ->    <-            E2 binding ->         <-
       putative AP-1 binding
7861 ctctatagtt ccttataatt
```

```
LOCUS       HPV34        7723 bp ds-DNA             VRL       04-OCT-1993
DEFINITION  Human papillomavirus type 34 (HPV-34), complete genome.
ACCESSION   X74476
SOURCE      Human papillomavirus type 34 DNA.
REFERENCE   1  (bases 1 to 7723)
  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 7723)
  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     HPV-34 belongs to a group primarily associated with orogenital
            lesions with low oncogenic potential.  HPV-34 was initially
            isolated and cloned from a squamous cell carcinoma of Bowen's
            type in 1986 by Kawashima et al. (J. Virol. 57:688-92)
            and subsequently has been sequenced by Dr. H. Delius.  It has
            also been detected in a genital intraepithelial
            neoplasia and periungual Bowen's disease. A study which probed
            lesions with Bowen's disease and squamous cell carcinomas for
            HPV-34 DNA, reported only one case of positive hybridization,
            indicating that HPV-34 infection of this nature is relatively
            rare (Kawashima et al. J Virol 57: 688-92 (1986)).

BASE COUNT     2475 a   1339 c   1614 g   2295 t
ORIGIN      101 bp upstream from beginning of E6 cds
        1 actataatcc tactataaaa tatagggtgT AACCGAAAAC GGTtgcaACC GATATCGGTg
                                 E6 orf start ->
                                     E2 bind ->       E2 bind ->
       61 catatataag tgctgcagta cactgctgga cagattggga aATGtttttt cccaatcctg
                                                 E6 cds ->
      121 aggaacggcc atacaagcta ccagccttat gtgaagaggt caacatttca atacatgaaa
      181 tagaattgga ctgtgtgtat tgcgaacgac aactgtacag atgtgaggta tatgatttta
      241 tatttagaga tttatgtgtt gtatatagaa aggggaaACC ACTTGGGGTa tgtcaaccgt
                                            E2 bind ->
      301 gtttactgtt ttactcaaag gttagacaat atagaagata taaccaatca gtgtatggac
      361 ggacgttaga gaatttaact aacaaacagt tgtgtaatat tttaataagg tgcggaaaat
      421 gccaaaaacc actgtgtcca ttggaaaagc aaaggcatgt agaTGAaaac aaacggtttc
                                                 E7 orf start ->
      481 accaaatagc ggatcagtgg accggacgct gtacacagtg ctggagacca tctgcaacag
      541 tggtgTAAga ATGcatggaa aaaaaccaag tgtgcaggac attgtgttag atctgaaacc
             E7 cds ->
               <- E6 end
      601 aacgaccgag actgACCTTA CATGTTacga gtcattagac aactcagagg atgaggatga
                E2 bind ->
      661 aacagacagc catctagaaa gacaagctga gcaagcctgg tacagaatag ttactgattg
      721 cagcagatgt cagtccacag tgtgtcttac cattgagagc acacacgctg acctattagt
      781 gttagaagac ctgcttatgg gtgcactaaa aattgtgtgc cccaactgtt ccagacgcct
      841 aTAAcagaag ATGgctgatt caggtaattg ggaagggagg tgttcgggat ggtttaatgt
             E1 cds ->
 E1 orf start ->
            <- E7 end
      901 agaagccatt gtagaaagga aaacagggga tgcaatacca gcagatgaaa attatgatgg
      961 ggacgataca gaggattctg aaatgggggа ttttattgat aatgcacaca tatctaatat
     1021 atattcacag caggaaattg cacaggcatt atatcactca cagcaagtta atgcagacaa
     1081 tgaggctata cgtgttctaa aacgaaagtt tgcaggtagt gctggcagta gcccagacag
     1141 taaaagacat gaattgaaac acaaacagcg tagtccacat atattgacga taagggacac
     1201 taatactaca tctacacacc tattgtgtga ggaacaagac agcggatatg gcaatactga
     1261 agtggaaacg tacgagagac aggtaccggg gccgggggga tgtttacaaa gtactagcag
     1321 tagtaacaac ggcagccaaa tggcgtcgcc aggggaaaca aatagtgggt ccagtagcat
     1381 ttcaaatatg gatatagaca tggaaagcac acctataacg gacattacaa acatattaaa
     1441 aagtagtaat gtaaaggcaa cattattagc aaaatttaaa gaggtatatg gattaagtta
     1501 tatggaatta gtaaggcctt ataaaagtga taaaacacag tgtcaggact gggtatgtgc
```

```
1561 agtgtttggg gtagcgccat cattggcaga aagtttaaaa tcattactaa cccaatattg
1621 cctatacata catctacaat gtttaacatg ttcgtggggt ataatagtat tattgttagc
1681 aagatttaag tgcaataaaa atagattgac agtacaaaaa ttattacatg ggttattaaa
1741 tgtaacacag gaatatatgt taatagaacc acctagacta agaagcacgc catgtgcatt
1801 atactggtac agaactagcc tatcaaacat tagtgaaacg gtgggagaag tacccgaatg
1861 gattaaaaga caaacagtag tacagcacag cttagaggac tgtcaatttg acctatctca
1921 aatggtacag tgggcatttg ataatgacat aacaaatgac tgtgaaatag catataaata
1981 tgcattatta gcatctgagg atagcaatgc tgctgcattt ttaaaaagca atgcacaagc
2041 aaaatatgtt aaggattgtg gaacaatgtg tagacattat aaagctgcag aacgtaaaca
2101 aatgactatg tcacaatgga ttacacatag atgtgattta atagatgatg gaggaaactg
2161 gaaacatatt gtgcaatttt taagatatca gcaggttgaa tttgtaccgt ttttaattgc
2221 tttaaaacaa ttttttaaagg gtataccaaa acaaaattgt atagttatat atggaccacc
2281 agatacagga aagtcacatt ttggaatgag tttaatgcag tttatgcaag gtgtggttat
2341 ttcatatgta aattccaata gtcatttttg gttatcgcca ttggctgatg caaaaatggc
2401 attattagat gatgcaacac ctgcatgctg gacatatatt gatagatatt taagaaatgc
2461 attagatggc aatcctatgt gtttagacag aaaacataaa catttattac aaattaaatg
2521 tcctccatta ctaataacat caaatacaaa tcctaaagca gatgatacct ggaaatattt
2581 gcacagtaga atgaaagtgt ttacgttttc aaatccattt ccatttgaca gtaatggaaa
2641 tccactatac caacttacta atgaaaactg gaaggcattt tttacaaaga cgtggtcaaa
2701 actagattTA Acagaggacg acgacaagga aaATGatgga gacactgtgc aaacgtttaa
    E2 orf start ->             E2 cds ->
2761 gtgcgtgtca ggacgcaatc ctagaactgt aTGAacgtga tagtatacat ttaagtgatc
                                       <- E1 end
2821 atattgatca ctggaaacac gtgcgactgg aaaatgtatt attacataag gcacgtgaaa
2881 tgggactgca atcagttaac caacaagcgg tgccaagcct tgcagtatca cgatccaaag
2941 ggcataatgc aattgaacta caattagccc tagaaagttt aaatgaatca agctataaca
3001 cagaagaatg gacattacaa cagacaagtt gggaacagtg ggtaacggac ccaaaacaat
3061 gttttaaaaa aggtggaaaa acagtagaag ttagatatga ctgtgacaag gacaacacca
3121 tgcaatatgt ggtatggaca tttgtgtatt attggttgga aggcaagtgg tataaagtga
3181 gtagccatgt agattataat ggtatatatt atgaaacaca ggacaatgaa aaggtatatt
3241 atacacaatt tgacagagat gcaaaacgat atggggttaa aggaatatgg gatgtatgta
3301 tgggcggtaa ggTAAtatgt tttgctcctg tatttagccc gtgtgaagta tccactcctg
       E4 orf start ->
       NH2 terminus unknown
3361 aaattgttag acccctgcac acaagcaaca gcagcaacgc acaggacgcg ggtgtgccaa
3421 cacggaaacg gcatagacag tgtgacccag acgaggggcc cttggacttt gtacataacc
3481 tacagcccac aacagactca tcgacccagt gtactctaca taatgttgcg ccaaTAGtac
                                                              <- E4 end
3541 atttaaaagg tgacaaaaac agtttaaaat gcttaagata taggatgcat aaagggtatt
3601 cacatttgtt taataatgta acaactacat ggcattggac caataataca aatagtaaat
3661 gtggtgtaat tacatttatg ttttccagta catcccaaca aaaacaattt ttacaatgtg
```

```
3721 ctaaaatacc accaactata tcagtgtcat cagggtatat gtccataTAA cacaaatatg
                                                        <- E2 end
                                                 E5 orf start ->
3781 taacattggt tacaattgta tacattggta tactgtataa tgtcATGatc atttttatat
                                                E5 cds ->
3841 ttgtaatttt gtttgtcttt ggcttttata tgtgcctgtc tgtgtcgctt gcagtgtctg
3901 tgtgtgttta tgcatggtta ttgcttttga ttataattac ttttttacat gtgtcccaat
3961 ccctgttaaa agtatatata ttatatgtgt gtgtgtttta tattcctatg gcactggtac
4021 attaccatgc taccttgcaa gtaacaTAAa tattttcata tatatacaat aaatttattg
                                    <- E5 end
4081 tttgtgtacc agtgtgTAGt ttaccATGcg tcgcaagcgt gacacacata taagacgtaa
                           L2 cds ->
            L2 orf start ->
4141 acgtgcatca gccacacaat tatacaaaac atgtaagcaa agtggcacat gtccccctga
4201 cataattcca aaggttgagg gcaatacttt agctgaccaa atATTAAAat atggtagcat
                                               signal ->
4261 tggggtattt tttgggggtt tgggtattgg tagcgggtca ggaaccgggg gtcgtacagg
4321 atatgtgcca ttacccacaa ctactccatc tagaccagtt gaaattcctt tgcaacctac
4381 cagacctcct gttattacat ctgtaggggc ctctgattct tctatagttt cattagtgga
4441 ggagtcaagt tttattgaag ctggtgtgcc aggtcctacc tccatagtgc cttctagttc
4501 tggatttaat gtaacaacat ctgtggacag tacacctgct attatagatg tggcaactat
4561 aagtgacact acacaagtat ctgttagtac atttaacaat ccaactttta ctgacccatc
4621 cgtgttgcaa cctcctccac ccttagaggc ctctggcaga cttttatttt caaatgatac
4681 tgtaactact cattcttatg aaaatatacc tcttgacaca tttgtagtta caacagataa
4741 caatagtatt gttagtagta cgcccattcc agggaggcac cctcctgcac gcttagggct
4801 ttatggacgt gctatacaac aggttaaggt tgtagaccct gcttttgtaa caacgcctac
4861 acgtttggta acatatgata atcctgcctt tgagggcctg caagatacca cattagaatt
4921 tcagcacagt gacttgcata atgctcctga ctctgatttt ttagatattg taaagttaca
4981 taggcctgct ttaactgcta gaaaaactgg catacgtgtt agtagactag gacaacgtgc
5041 cactatgttt actagaagtg gcaaacgtat aggtggtagg gtacattttt atcatgattt
5101 aagtccaata cctacggaaa atattgaact gcagccttta ctgccttctg catctgctac
5161 tgtaacagat gctaatggca ttaatgatgg gctatatgat gtattgttag ataataatgt
5221 agatattact gaagtggaaa cacctactgg tacaaacaca caaagtgttt ttgcaagtga
5281 gatttccact acaactgcaa ataccactat tccttTAAAT Gctggtttgg atacacatcc
                                 L1 orf and cds ->
5341 tggccctgat attgctttac ctgtacctac tgcagaaacc atttttaccc caactgtacc
5401 tgtacagcct tctggtccta tatatatata tgggtccgat tttatattgc atcctagctt
5461 gtatgtcatt ccacgcaaac gtaaacgttt gtcatatttt tttgcagatg tggcgaccta
5521 cTGAggcaaa ggtatacctg ccccctgtgt ctgtatctaa agttgtaagc actgaagagt
       <- L2 end
5581 atgtaacacg cacaaatata tattattatg caggtagtac acgcttgctg gcagtaggac
5641 atccctatta tcctataaag gatactaatg ggaaacgtaa gattgctgta cctaaagttt
5701 caggtttgca atacagggta tttagaatac gtttgccaga tcccaataaa tttgggttcc
5761 cagatgcatc tttttataat cctgacaagg agcggttggt ttgggcctgt gcaggcgttg
5821 aggttggacg cggacagcca ttaggtatag gtactagtgg taatcctttt atgaataaac
5881 ttgaggatac tgaaaatgct gcaaaatata ttggtggaaa catagcagat agtagggagt
5941 gtatgtcagt agattataaa caaacacagt tatgtattgt aggctgtaag cctcccttag
6001 gggaacattg gggtacaggc acaccatgtg gtacccaaaa tgctggggaa tgtcctccgc
6061 ttgaattaaa gaacaccact atacaggatg gtgatatgat tgatgtgggc tttggtgcca
6121 tggattttaa ggccttacag gcaAATAAAa gtgatgtgcc cattgatatt tccaacacta
                       signal ->
6181 tatgtaaata tccagattat ctaggcatgg ccgcagatcc ctatggcgat tctatgtggt
6241 tttatattcg tagggaacaa atgtttgtta gacacctatt taacagggca ggtactgtag
6301 gcgatgctat tccagatgac ttaatgatta agggtacagg caatactgca tcgccatcca
6361 gttgtgtgtt ttatcctact cctagtgggt ctatggtatc ctcagatgca caaattttta
6421 ataagcctta ttggttgcaa aaggcccagg gacaaaacaa tggcatttgc tggcataatc
6481 aactgttttt aactgttgta gatactacta gaagcacaaa cttttcagtt tgtgtaggta
6541 cacaatccac aagtacaact gcaccatatg caaacagtaa ttttaaggaa tacctcagac
6601 atgcagaaga gtatgacctg cagtttgtgt ttcagttatg caaaatcaat ttaactacag
6661 atgtaatgac atatatacat tctatgagtt ctagtatatt ggaacagtgg aattttggtc
6721 ttacaccacc gccttcaggt actttagaag aaacatatag atatgttact tcacaggcca
6781 ttacatgtca gcgtccgcaa cctcctaagg aaacagagga cccatatggt aaaatgacat
6841 tttgggaggt agaccttaaa gaaaaatttt ctgcagaatt agatcagttt gccttaggta
6901 gaaagttttt attacaacta ggtatgcgtg cccgtcctag gttacaggcc tctaaacggt
```

HPV34

```
6961 ctgcaccttc atcatctagt acagcaccta agaaaaaacg tgcaaaacgt attTAAttaa
                                                              <- L1 end
7021 atgtatgtgt gttgtgtttg tgtatgtgta ttgtgtgttt gtatatgttt gtgtatgtgt
7081 attgtgtgtt tgtatatgtt atgtaatgta tgtattattg tattgttgta tgtggtgtgg
7141 tattgtatgg catatggtgt tgtacatgtg tgtgtatgta tgtatgtatg tgttttatgt
7201 tcaataaact atatatgttg tgtgagtggt ttttaatagc cactgtgtct ccattttgtt
7261 cattcgccat tttatatatt ttaaagttat tctgcgACCG CTTTCGGTtg cctatactat
                                  E2 bind ->
7321 aacatgtttt ggtctgtttg tgctgtttta gcacaaaggt ttaatgtgtt ttggcttcct
7381 gcaggcaact tggattgtac tcaagcactt aggattaaca tgcttcctgc ttatgtttca
7441 tcctacgcaa ggtcatacac gtttaaggcg ccacctggca gttactcatg tgtctgcaat
7501 tgttactaca atgttttgca cacacatttt ttacccaccc tctcataaaa ttgcttttaa
7561 tcacatactt gtactatgta aacagtgtac cttggcagaa cattgtattt tacatgccag
7621 gtaattgttg aaaactgact aagcagtgct tactcatttt gcACCTGGAC TGTTaaaccg
                                       E2 bind ->
7681 ttttggatca cacagtgtac caacccttt ataattataa aaa
```

```
LOCUS       HPV44E6       590 bp ds-DNA             VRL       15-SEP-1989
DEFINITION  Human papillomavirus type 44 (HPV-44), E6 region.
ACCESSION   M27023
SOURCE      Human papillomavirus type 44 DNA recovered from a vulvar
            condyloma.
REFERENCE   1  (bases 1 to 590)
  AUTHORS   Lorincz,A.T., Quinn,A.P., Goldsborough,M.D., Schmidt,B.J. and
            Temple,G.F.
  TITLE     Cloning and partial DNA sequencing of two new human papillomavirus
            types associated with condylomas and low-grade cervical neoplasia
  JOURNAL   J. Virol. 63, 2829-2834 (1989)
COMMENT     HPV-44 is a mucosatropic HPV which to date has not been detected in
            cervical cancer.  Prevalence studies indicate that HPV-44 and HPV-43
            have been found in 4% of cervical intraepithelial neoplasms, but in
            none of the 56 cervical cancers tested.

            During the analysis of approximately 1000 anogenital tissue samples,
            two new HPV types, HPV-43 and HPV-44, were identified.  The complete
            genome of HPV-44 was recovered from a vulvar condyloma and cloned
            into bacteriophage lambda.  The biopsy was taken from a woman from
            the Detroit Michigan area.  The DNA recovered was a single 7.8 kb
            BamHI fragment.  Only the E6 region of the cloned sample has been
            sequenced, although the positions of the ORFs for the entire genome
            have been deduced and are consistent with the organization of DNA
            from HPV-6b.  A possible feature of HPV types associated with
            malignant lesions is the potential to produce a different E6 protein
            by alternative splicing.  This potential has been found in types
            HPV-16, HPV-18, and HPV-31.  HPV-44 has a potential E6 splice donor
            at nt 229, but does not contain a potential splice acceptor.

BASE COUNT      197 a    117 c    124 g    152 t
ORIGIN      102 bp upstream from beginning of E6 cds
        1 aataataatc taacctttac aaaaaagagg aggaACCGAA TTCGGTtcca ACCGAAAACG
                                       E2 bind ->        E2 bind ->
       61 GTtataTAAa aaccagccca aaaattaagc aagcggggca taATGgaaag tgcaaatgcc
      E6 orf start ->                                E6 cds ->
      121 tccacgtctg cacaaagtat agaccagttg tgcaaggagt gcaacattcc tatgcacaat
      181 ctgcaaattt tatgcgtgtt ttgcagaaaa acgttaagta ctgcagaggt ttattcattc
      241 gcatataaac agttatatgt agtgtaccga ggaaactttc catttgcagc ctgtgccatt
      301 tgtttagaac tacaaggtaa ggtcaatcaa tttaggcatt ttaactacgc gggatatgca
      361 gtaacagtgg aagaagaaac aaataagtca attctggacg tgctgatacg ctgctatttg
      421 tgccacaaac cattgtgcca cgtggaaaag gtgcgccaca tattggacaa ggcgcgattc
      481 attaaattac aagatACCTG GAAGGGTcgc tgcttccatt gttggacatc atgcatggaa
                  E2 bind  ->
      541 actatactac ctTAAaggaa attgttttac agctggaacc tcctgaccct
                        <- E6 end
//
```

```
LOCUS       HPV44MY911    455 bp ds-DNA             VRL       16-OCT-1994
DEFINITION  Human papillomavirus type 44 (HPV-44), partial L1 cds, My09/My11
            region.
ACCESSION   U12493
SOURCE      Human papillomavirus type 44 DNA recovered from a paitient
            with a vulvar condyloma.
REFERENCE   1  (bases 1 to 455)
  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-44 DNA was first isolated from a vulvar condyloma.
            The cloned DNA was provided by Dr. A. Lorincz and was subsequently
            sequenced by Dr. H. Delius over the L1 MY09/MY11 segment.  HPV-43
            and the several other types sequenced by Dr. Delius over this region
            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      135 a     90 c    100 g    130 t
ORIGIN
        1 gcgcagggcc acaataatgg tatttgttgg ggaaatcagt tatttgttac tgttgtagat
  L1 cds ->
          -> MY11 PCR primer <-
       61 actacccgta gtacaaacat gacaatatgt gctgccacta cacagtcccc tccgtctaca
      121 tatactagtg aacaatataa gcaatacatg cgacatgttg aggagtttga cttacaattt
      181 atgtttcaat tatgtagtat taccttaacg gcggaggtaa tggcctatct tcatactatg
      241 aatgctggta ttttagaaca gtggaacttt gggttgtcgc cgcccccaaa tggtacctta
      301 gaggacaaat acagatatgt gcagtcccag gccattacat gtcaaaagcc acccccctgaa
      361 aaggcaaagc aggaccccta tgcaaaatta agtttttggg aggtggatct tagagaaaag
      421 ttttctagtg agttggatca atatcccctt ggtag
                                        L1 cds ->
                          -> MY09 PCR primer  <-
```

```
LOCUS       HPV55MY911    455 bp ds-DNA             VRL       16-OCT-1994
DEFINITION  Human papillomavirus type 55 (HPV-55), partial L1 cds, My09/My11
            region.
ACCESSION   U12494
SOURCE      Human papillomavirus type 55 DNA.
REFERENCE   1  (bases 1 to 455)
  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-55 was first isolated from a penile condyloma acuminata.
            Cloned HPV-55 DNA was obtained from the Papillomavirus Reference
            Center, Heidelberg and subsequently sequenced by Dr. H. Delius
            over the L1 MY09/MY11 segment.  HPV-55 and the several other HPV
            types recently sequenced over the MY09/MY11 primer 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      143 a     86 c     94 g    132 t
ORIGIN
        1 gcgcagggcc acaataatgg tatttgttgg gggaatcagt tatttgttac tgttgtagat
  L1 cds ->
          -> MY11 PCR primer <-
       61 actacacgta gtacaaacat gacaatatgt gctgctacaa ctcagtctcc atctacaaca
      121 tataatagta cagaatataa acaatacatg cgacatgttg aggagtttga cttacagttt
      181 atgtttcaat tatgtagtat taccttaact gctgaggtaa tggcctattt acataccatg
      241 aatcctggta ttttggaaca gtggaacttt gggttgtcgc cacccccaaa tggtacctta
      301 gaagacaaat acagatatgt gcagtcacag gccattacat gtcaaaagcc tccccctgaa
      361 aaggcaaagc aggaccccta tgcaaaatta agtttttggg aggtagatct cagagaaaag
      421 ttttctagtg agttagatca atatcccctt ggtag
                                        L1 cds ->
                          -> MY09 PCR primer  <-
```

```
LOCUS       HPV64MY911    458 bp ds-DNA             VRL       16-OCT-1994
DEFINITION  Human papillomavirus type 64 (HPV-64), partial L1 cds, My09/My11
            region.
ACCESSION   U12495
SOURCE      Human papillomavirus type 64 DNA recovered from a patient with
            vulvar intraepithelial neoplasia (VaIN).
REFERENCE   1  (bases 1 to 458)
  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-64 has recently been characterized and isolated from a vulvar
            intraepithelial neoplasia by Dr. T. Matsukura.  The cloned DNA
            was subsequently sequenced by Dr. H. Delius over the L1 MY09/MY11
            segment.  HPV-64 and the several other 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      152 a     85 c     83 g    138 t
ORIGIN
        1 gcacagggac ataacaatgg aatttgttgg cataatcaac tgtttctaac tgttgtatat
  L1 cds ->
          -> MY11 PCR primer <-
       61 actaccagaa gtacaaactt ttctgtttgt gtaggcacac aatccacaag tacaaatcca
      121 ccatatgcaa acactaattt taaggaatac ctaaggcatg cagaagagta tgacctcgag
      181 tttgtgtttc agttatgcaa aattacttta actacagatg taatgacata tatacattct
      241 atgagttcta gtatattaga acagtggaat tttggtctta caccaccgcc atctggtact
      301 ttagaagaaa catatagata tgtaacttca caggccatta catgtcagcg tccgcaacct
      361 cctaaggaat ctgaggatcc atatgctaaa atgacatttt gggaggtaga ccttaaagaa
      421 aagttttctg cagaattaga tcagtttcct tttggacg
                                           L1 cds ->
                            -> MY09 PCR primer  <-
```

```
LOCUS       HPVMM9        458 bp ds-DNA             VRL       16-OCT-1994
DEFINITION  Human papillomavirus isolate MM9, partial L1 cds, My09/My11
            region.
ACCESSION   U12491
SOURCE      Human papillomavirus DNA recovered from a genital swab sample,
            isolate MM9 (PAP238a).
REFERENCE   1  (bases 1 to 458)
  AUTHORS   Manos,M.M., Waldman,J., Zhang,T. Greer,C., Eichinger,G.,
            Schiffmann,M., and Wheeler, C.
  TITLE     Epidemiology and partial nucleotide sequence of four novel genital
            human papillomaviruses
  JOURNAL   J. Infect. Dis. (1994) In press
COMMENT     MM9, also known as PAP238a, was isolated from a genital swab sample.
            Samples were obtained from over 500 patients examined at either the
            Shasta/Diablo planned parenthood clinic or at a private practice
            in the state of California over the course of seventeen months.
            Each of the samples were cervical or vulvar/intraoital in origin.
            DNA was PCR amplified over the MY09/MY11 region and subsequently
            sequenced if the HPV digested products yielded unique RFLP patterns.
            This procedure resulted in the identification of four novel HPV
            types: W13B, PAP291, PAP155, and PAP238a, which have subsequently
            been renamed MM4, MM7, MM8, and MM9. Oligonucleotide probes over
            the MY9/MY11 region from these viruses have been reported by
            Hildesheim et al. (J Infect Dis 169: 235-40).   These probes were
            used to determine prevalence in different populations.  Prevalence
            for each of these viruses was similar to that seen in other
            characterized "intermediate risk" viruses probed for in these
            studies. In addition to anogenital sites, MM9 was identified in a
            periungual carcinoma with multiply infected digits.  It should be
            noted that MM4 is extremely similar (90.8%) to novel HPVIS39 (U12481)
            and MM7 is virtually identical to LVX82 (U12487).  Primer regions are
            annotated in the sequence; information in this region is not accurate
            due to primer degeneracy.

BASE COUNT      145 a     79 c     84 g    150 t
ORIGIN
        1 gcacagggtc ataataatgg tatttgttgg cataatcaat tatttttaac tgttgtagat
  L1 cds ->
          -> MY11 PCR primer <-
       61 actactagaa gcactaattt ttctgtatgt gtaggtacac aggctagtag ctctactaca
      121 acgtatgcca actctaattt taaggaatat ttaagacatg cagaagagtt tgatttacag
      181 tttgttcttc agttatgtaa aattagttta actactgagg taatgacata tatacattct
      241 atgaattcta ctatattgga agagtggaat tttggtctta ccccaccacc gtcaggtact
      301 ttagaggaaa catatagata tgtaacatca catgctatta gttgccaacg tcctcaacct
      361 cctaaagaaa cagaggaccc atatgccaag ctatcctttt gggatgtaga tcttaaggaa
      421 aagttttctg cagaattaga tcagtatccc cttggacg
                                           L1 cds ->
                            -> MY09 PCR primer  <-
```