Group H Sequences

HPV5   **HPV8**
HPV9   **HPV12**
HPV14d   **HPV15**
HPV17   **HPV19**
**HPV20**   **HPV21**
HPV25   **HPV47**
HPV49

INTRODUCTION

Group H consists of human papillomaviruses HPV-5, HPV-8, HPV-9, HPV-12, HPV-14d, HPV-15, HPV-17, HPV-19, HPV-20, HPV-21, HPV-25, HPV-47, and HPV-49, a group primarily associated with the multifactorial disease, Epidermodysplasia Verruciformis (EV). Patients with EV tend to have depressed cell mediated immunity [1]. In roughly one-third of EV-associated HPV infection, the sun-exposed flat wart-like or macular lesions transform into malignant squamous cell carcinomas [2]. Benign wart scrapings tend to be multiply-infected, with as many as six different viral types. However, in contrast, EV carcinomas tend to harbor only a few types, specifically HPV-5 and HPV-8, and less frequently HPV-14, HPV-17, HPV-20 and HPV-47 [2, 3]. These types are rarely detected in lesions afflicting the general population.

A key to host restriction of these viruses may be in part due to the unusual organization of the LCR in these viruses. The LCR of these viruses is short compared to the viruses in other groups and contains two EV-specific regulatory regions: M33 and M29, both shown to be involved in protein binding [3, 4, 5].

This group forms two major branches based on phylogenetic analysis, one which can be subdivided into two minor branches. These clusters have been designated as a_1, a_2, and b. In addition, HPV-49 forms a remote branch off of the b cluster. HPV-49 is curious in so far as EV associated lesions; the awkwardness of its position is perhaps reflected in the distance of its relationship to the other papillomaviruses in the group.

Ensser proposed a classification scheme of those sequenced EV types based on the presence or absence of conserved EV-specific and other regulatory regions within the LCR. This categorization is consistent with that obtained through our phylogenetic analysis. In this system, viruses in groups A1 and A2 possess the EV-specific M33 and M29 regulatory regions, however viruses in the B group contained only segments of these motifs. Subgroup A2 differed from A1 by the presence of two of four degenerate E2 binding sites [4]. Other researchers have also devised classification schemes based on other criteria [6]. Dr. H. Pfister classified the EV-asociated viruses by the level of cross-hybridization to each other and to those in other groups. In his system, the D1 viruses correspond to both the a_1 and a_2 categories proposed here. D2 corresponds to the b cluster, and D3 is composed solely of HPV-24, a virus not presented in this compendium [7].

Cluster a_1 consists of HPV-5, HPV-8, HPV-47 and HPV-12. Both HPV-5 and HPV-8 are associated with macular lesions which frequently progress to malignancy [8, 9, 10]. Yabe et al. studied the characteristics of HPV-5 in lesions of differing severity. In a primary carcinoma, HPV-5 was present in an episomal state with a 40% subgenomic segment amplified. In the metastatic tumor, only the 40% subgenomic region was present, but integrated into the host genome [10]. The segment was determined to be the entire sequences of E6, E7, and the noncoding region and portions of E1 and L1, with no mutations present [11]. In addition, amplifications of the LCR have been reported in
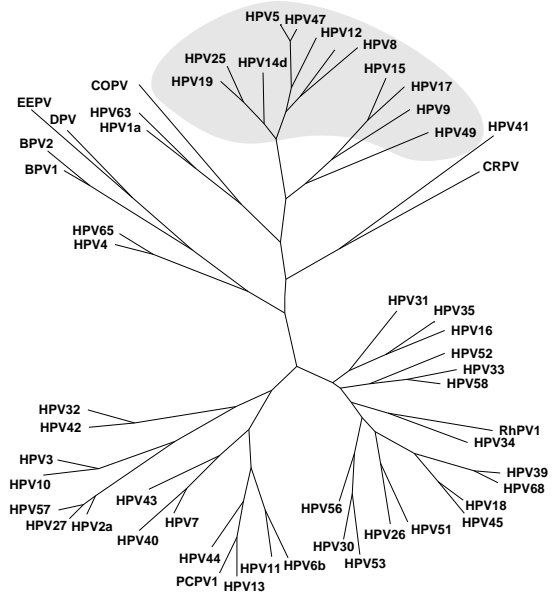
HPV-5 associated carcinomas [12]. HPV-5 and HPV-8 have also been found in significant numbers in squamous cell carcinomas of renal allograft patients. Barr et al. detected either HPV-5 or HPV-8 in nearly 60% of the cases surveyed in Scotland [13]. HPV-47 is primarily associated with benign lesions, however, it has also been detected in cases of malignancy [6]. HPV-12 induces benign macular and flat wart-like lesions [14].

Cluster a$_2$ consists of HPV-19, HPV-25, HPV-14, HPV-21 and HPV-20. HPV types forming this cluster produce benign macular or flat wart-like lesions and malignant lesions in isolated cases. Both HPV-19 and HPV-25 induce macular lesions, which are benign in character [7, 6, 15]. HPV-14, HPV-20 and HPV-21 induce flat-wartlike lesions; HPV-20 and HPV-14 have been detected in carcinomas [6, 15].

Cluster b consists of HPV-15, HPV-17, and HPV-9. HPV-15 was isolated from a benign flat wart-like lesion [15]. HPV-17 was isolated from benign macules and subsequently from squamous cell carcinomas and the malignant melanoma of an immunosuppressed patient [15, 16]. HPV 9 DNA induces both macular and flat wart-like lesions, however it has also been identified in a keratoacanthoma [14, 17].

HPV-49, a type which clusters with the b group of the EV associated viruses, was isolated from the flat warts of a Polish renal transplant patient. Favre et al. screened benign and malignant lesions from the general population, EV patients and transplant patients for the presence of HPV-49. In the survey, HPV-49 was not detected in any of the patients with EV but was detected in two additional cases of flat warts in renal transplant patients [18].

HPV-5 and HPV-47 are close enough to each other to be considered "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 these changes are "silent", causing no difference at the amino acid level (Part III).

---

[1] Lutzner,M.A. Epidermodysplasia verruciformis. An autosomal recessive disease characterized by viral warts and skin cancer. A model for viral oncogenesis. *Bull Cancer (Paris)* **65**: 169–82 (1978)

[2] Gassenmaier,A., Lammel,M., and Pfister,H. Molecular cloning and characterization of the DNAs of human papillomaviruses 19, 20, and 25 from a patient with epidermodysplasia verruciformis. *J Virol* **52**: 1019–23 (1984)

[3] May,M., Helbl,V., Pfister,H., Fuchs, P.G. Unique topography of DNA-protein interactions in the non-coding region of epidermodysplasia verruciformis-associated human papillomaviruses. *J Gen Virol* **72** (Pt 12): 2987–97 (1991)

[4] Ensser,A. and Pfister,H. Epidermodysplasia verruciformis associated human papillomaviruses present a subgenus-specific organization of the regulatory genome region. *Nucleic Acids Res* **18**: 3919–22 (1990)

[5] Krubke,J., Kraus,J., Delius,H., Chow,L., Broker,T., Iftner,T., and Pfister, H. Genetic relationship among human papillomaviruses associated with benign and malignant tumours of patients with epidermodysplasia verruciformis. *J Gen Virol* **68** (Pt 12): 3091–103 (1987)

[6] Kiyono,T., Hiraiwa,A., and Ishibashi,M. Differences in transforming activity and coded amino acid sequence among E6 genes of several papillomaviruses associated with epidermodysplasia verruciformis. *Virology* **186**: 628–39 (1992)

[7] Pfister, H. Papilomaviruses: general description, taxonomy, and classification. In *The Papovaviridae 2* (N. P. Salzman and P.M. Howley, Eds.), pp. 1–38. Plenum, N.Y.

[8] Zachow,K.R., Ostrow,R.S. and Faras,A.J. Nucleotide sequence and genome organization of human papillomavirus type 5. *Virology* **158**: 251–4 (1987)

[9] Orth,G. Epidermodysplasia verruciformis: a model for understanding the oncogenicity of human papillomaviruses. *Ciba Found Symp* **120**: 157–74 (1986)

[10] Yabe,Y., Tanimura,Y., Sakai,A., Hitsumoto,T., and Nohara, N. Molecular characteristics and physical state of human papillomavirus DNA change with progressing malignancy: studies in a patient with epidermodysplasia verruciformis. *Int J Cancer* **43**: 1022–8 (1989)

[11] Yabe,Y., Sakai,A., Hitsumoto,T., Kato,H., and Ogura,H. A subtype of human papillomavirus 5 (HPV-5b) and its subgenomic segment amplified in a carcinoma: nucleotide sequences and genomic organizations. *Virology* **183**: 793–8 (1991)

[12] Deau,M.C., Favre,M., and Orth, G. Genetic heterogeneity among human papillomaviruses (HPV) associated with epidermodysplasia verruciformis: evidence for multiple allelic forms of HPV5 and HPV8 E6 genes. *Virology* **184**: 492–503 (1991)

[13] Barr,B.B., Benton,E.C., McLaren,K., Bunney,M.H., Smith,I.W., Blessing,K., Hunter,J.A. Human papilloma virus infection and skin cancer in renal allograft recipients. *Lancet* **1**: 124–9 (1989)

[14] Kremsdorf,D., Jablonska,S., Favre,M., Orth,G. Human papillomaviruses associated with epidermodysplasia verruciformis. II. Molecular cloning and biochemical characterization of human papillomavirus 3a, 8, 10, and 12 genomes. *J Virol* **48**: 340–51 (1983)

[15] Kremsdorf,D., Favre,M., Jablonska,S., Obalek,S., Rueda,L.A., Lutzner,M.A., Blanchet-Bardon,C., Van Voorst Vader,P.C. and Orth,G. Molecular cloning and characterization of the genomes of nine newly recognized human papillomavirus types associated with epidermodysplasia verruciformis. *J Virol* **52**: 1013–8 (1984)

[16] Yutsudo,M and Hakura,A. Human papillomavirus type 17 transcripts expressed in skin carcinoma tissue of a patient with epidermodysplasia verruciformis. *Int J Cancer* **39**: 586–9 (1987)

[17] Scheurlen,W., Gissmann,L., Gross,G., and zur Hausen,H. Molecular cloning of two new HPV types (HPV 37 and HPV 38) from a keratoacanthoma and a malignant melanoma. *Int J Cancer* **37**: 505–10 (1986)

[18] Favre,M., Obalek,S., Jablonska,S., and Orth,G. Human papillomavirus type 49, a type isolated from flat warts of renal transplant patients. *J Virol* **63**: 4909 (1989)

```
LOCUS       HPV5            7746 bp ds-DNA            VRL        30-SEP-1988
DEFINITION  Human papillomavirus type 5 (HPV-5), complete genome.
ACCESSION   M17463
KEYWORDS    complete genome.
SOURCE      Human papillomavirus type 5 DNA recovered from a benign flat
            wart from an EV patient.
REFERENCE   1  (bases 1 to 7746)
  AUTHORS   Zachow,K.R., Ostrow,R.S. and Faras,A.J.
  TITLE     Nucleotide sequence and genome organization of human papillomavirus
            type 5
  JOURNAL   Virology 158, 251-254 (1987)
COMMENT     Draft entry and printed copy of sequence for [1] kindly provided by
            R.S.Ostrow, 10/23/87.

            HPV-5 has been associated with macular lesions which frequently
            progress to malignancy.  Yabe et al. (Int J Cancer 43: 1022-8)
            studied the characteristics of HPV-5 in lesions of differing
            severity.  In a primary carcinoma, HPV-5 was present in an episomal
            state with a 40% subgenomic segment amplified.  In the metastatic
            tumor, only the 40% subgenomic region was present, but integrated
            into the host genome.  The segment was determined to be the entire
            sequences of E6, E7, and the noncoding region and portions of E1 and
            L1, with no mutations present (Yabe et al. Virology 183: 793-8).
            In addition, amplifications of the LCR have been reported in HPV-5
            associated carcinomas (Deau et al. Virology 184: 492-503).  HPV-5
            and HPV-8 have also been found in significant numbers in squamous
            cell carcinomas of renal allograft patients.  Barr et al. (Lancet
            1: 124-9) detected either HPV-5 or HPV-8 in nearly 60% of the cases
            surveyed in the Scotland area.  HPV-5 is considered to be part of
            the a$_1$ cluster based on phylogenetic analysis.  This cluster
            includes HPV-5, HPV-8, HPV-47, and HPV-12.  Patients with EV tend to
            have depressed cell mediated immunity.  In roughly one-third of
            EV-associated HPV infection, the sun-exposed flat wart-like or
            macular lesions transform into malignant squamous cell carcinomas.
            Benign wart scrapings tend to be multiply-infected, with as many as
            six different viral types.  However, in contrast, EV carcinomas tend
            to harbor only a few types, specifically HPV-5 and HPV-8, and less
            frequently HPV-14, HPV-17, HPV-20 and HPV-47.  These types are
            rarely detected in lesions afflicting the general population.   A
            key to host restriction of these viruses may be in part due to the
            unusual organization of the LCR in these viruses. The LCR of these
            viruses is short compared to the viruses in other groups and
            contains two EV-specific regulatory regions: M33 and M29, both
            shown to be involved in protein binding.

BASE COUNT     2376 a   1547 c   1736 g   2087 t
ORIGIN      354 bp upstream of HindIII site.
        1 AACGGTaagt tgcaatttcc ttgtaccagg tgcggtattg ggatttcaca atTATAATgg
   E2-bind <-                                          signal ->
       61 ttgttgccaa ctaccatagg catattcaag tttttgcctg tatcgttttc gtatcctgta
      121 ataatatcca atatatgtat acataAATAA ATATATATAT ATATAAgtgt ctaagattgg
                           signal ->        E6 orf start ->
                                signal ->
      181 gttcttctgt aatcaggcaA TGgctgaggg agccgaacac caacagaaac tgacagaaaa
                  E6 cds ->
      241 agataaggca gaattacctt taagtattag agacttagct gaagccttag gcatccctgt
      301 gattgattgt ttaatacctt gcaatttctg tggcaacttt ctaaattatt tggaagcttg
      361 tgaattcgac tacaaaaggc ttagtctaat ttggaaagat tattgtgtgt ttgcgtgctg
      421 tcgcgtatgc tgtggcgcca ctgcaactta tgaatttaac caattttatg agcagacagt
      481 gttaggaaga gatattgaat tagcttcagg actttcaata tttgatattg atatcaggtg
      541 tcaaacttgc ttagcatttc ttgacattat agaaaagtta gattgctgtg gcagaggcct
      601 tcccttttcat aaggTGAgga acgcctggaa gggaatcgtg aggcagtgta agcatttta
                 E7 orf start ->
      661 tcATGattgg TAAagagagtc accgtgcaag atattattct ggagctcagt gaggtgcagc
   E7 cds ->         <- E6 end
```

```
 721 ccgaagtgct accagttgac ctgtttttgtg aagaggaatt accaaacgag caggaaacgg
 781 aggaggagcc tgacaacgaa aggatctctt acaaagttat agctccgtgc ggttgcagga
 841 actgtgaggt caagcttcgc attttttgtcc acgccacaga atttggtatt agagctttcc
 901 aacagctacT GAccggagat ctgcagctcc tgtgccctga ctgtcgcgga aactgcaaac
     E1 orf start ->
 961 ATGacggatc cTAAttctaa aggtagtaca tctaaagaag ggtttggtga ttggtgttta
E1 cds ->           <- E7 end
1021 ttggaagctg actgtagtga tgtagaaaat gatttgggac aattatttga gagagataca
1081 gactctgata tatcggattt gttagatgat actgaactgg agcagggcaa ttccctggaa
1141 ctatttcatc aacaggagtg tgagcagagc gaggagcaat tgcaaaaact aaaaacgaaag
1201 tatcttagtc caaaagctgt cgcacagctt agtccgcgac ttgagtcaat ttcattgtca
1261 ccccagcaga agtctaagcg caaggctcttt gcagagcagg acagcggact cgagctgact
1321 ttaaacaatg aagctgaaga tgttactcct gaggtggagg taccggctat tgactctcgg
1381 ccggatgacg agggaggttc aggggacgta gatatacatt acactgcatt gttgcgttct
1441 agcaacaaaa aagctacatt aatggctaag tttaaagagt cgtttggagt aggttttaat
1501 gaattgacac ggcaattcaa aagccacaaa acctgctgta aggactgggt tgtctctgta
1561 tatgcagtgc atgatgatct atttgaaagc tcaaagcagc tattgcaaca gcattgtgac
1621 tatatctggg tccgtgggat aggtgcaatg tcattatACC TATTGTGTTt taaggcggga
                                    -> E2 bind
1681 aaaaatcgcg ggacagttca taagttaatt acctcaatgt taaatgtgca tgaacagcaa
1741 atattgtctg agccgccaaa attgagaaat acagccgctg cattgttctg gtataagggt
1801 tgtatgggat cggggggcgtt tagccatgga ccatatcctg attggattgc ccaacaaact
1861 atattaggtc acaaaagtgc tgaggcaagt acttttgatt tttcagcaat ggtccaatgg
1921 gcatttcata atcacttatt agacgaagca gatatagcat accagtatgc aaggcttgct
1981 cccgaagacg cgaatgcagt agcttggctt gcacataaca ccaggccaa atttgtgaga
2041 gaatgtgcat atatggtacg attttataag aagggacaaa tgagagacat gagtatatct
2101 gaatggatat acactaaaat caatgaagta gaaggggaag ggcactggtc agatatagta
2161 aagtttatta gataccaaaa tataaacttt attgtattcc taactgcatt aaaagaattc
2221 ctacactcag tgccaaaaaa aaattgcatt ttaatttatg gtcctccaaa ttctggaaag
2281 tcatcatttg caatgtcatt aataagagtg ttgaagggta gagtgttgtc atttgtaaat
2341 tctaaaagtc agttttggct gcaacccctt tcagagtgca agatagctct attggatgat
2401 gtaacagacc cttgttggat atacatggat acatatttaa gaaatggctt ggatggacat
2461 tatgtttcat tagattgtaa atatagagcc ccaacgcaaa tgaaatttcc cccattatta
2521 ttaacatcta acattaatgt gcatggggaa actaattata gatatttaca cactacaata
2581 aaaggatttg aatttccaaa tcctttttcct atgaaagcag ataatacacc tcagttcgaa
2641 ctaactgacc aaagctggaa atctttttttt acaaggcttt ggacacaatt agaccTGAgt
                                                            E2 orf start ->
2701 gatcaagaag aggagggcga ggATGgagaa tctcagcgag cgtttcaatg ctctgcaaga
                E2 cds ->
2761 tcagctaatg aacatttaTG Aagctgcaga acaaacattg caggcacaaa ttaaacattg
                <- E1 end
2821 gcaaaccttta cgaaaagaac ctgtattact ctactatgct agggagaaag gtgttacaag
2881 gcttggatat caacctgtgc ctgtaaaggc agtatcagaa acaaaggcta aagaagccat
2941 agcaatggtc ctgcagcttg agtcactaca gacatctgat tttgctcatg agccatggac
3001 tctagttgat accagcatag aaacatttag aagcgctcca gaaggtcact tcaaaaaagg
3061 cccccctccct gtagaagtta tttatgacaa tgatccagat aatgccaatt gtatacaat
3121 gtggacctat gtgtattata tggatgcgga tgataagtgg cataaggcaa gaagtgggggt
3181 gaatcacatt ggcatttatt atttacaagg aactttttaaa aactattatg tactgtttgc
3241 tgacgatgcg aaaagatatg gtacaactgg agaatgggaa gTAAaagtta ataaggaaac
                                    E4 orf start ->
                                    NH2 terminus unknown
3301 tgtgtttgct cctgtcacca gctccacgcc tccagggtcg ccaggaggac aagcagacac
3361 aaacaccacc cccgcgaccc ccaccacctc cacaaccgcc gtTGActcca cgtccagaca
                                    E5 orf start ->
                                    NH2 terminus unknown
3421 gctcaccaca tcaaaacagc cacaacaaac cgaaaccaga ggaagaaggt acggacggag
3481 gccctccagc aagtcaagga gatcgcaaac gcagcaaagg cgatcaaggt cccgacACCG
                                                            E2-bind ->
3541 GTCCCGGTct cggtcccggt cgcggtccaa gtcccaaacc cacaccactc ggtccaccac
3601 caggtcccgg tccacgtcgc tcaccaagac tcgggcccttt acaagcagat cgcgatccag
3661 aggaaggtcc caaccacct gcagaagggg agtggaagg tcacccaggc ggcgatcaag
3721 gtcaccctcc acctcctcct cctgcaccac acaacggtca cagcgggcac gagccgaaag
3781 ttcaacaacc agagggggccc gagggtcgag agggtcacga ggagggagcc gtggggggag
3841 agggcggcga cgaggaaggt catcctcctc ctcctcccccc gcccacaaac ggtcacgagg
```

```
3901 ggggtctgcT AAgctccgtg gcgtctctcc tggtgaagtg ggagggtcac ttcgatcagt
                 <- E5 end
3961 tagttcaaag catacaggac gacttggaag attactggaa gaagctcgcg accccccagT
4021 AAtcattgtc aaaggggcgg ctaacacact gaaaaatgtc cgcaacagag ctaaaattaa
     <- E4 end
4081 atacatggga ctgtttaggt catttagtac tacctggtca tgggtggcag gagatggcac
4141 tgagcgtcta ggcaggccca gaatgctcat tagcttttct tcctatactc aaaggagaga
4201 ttttgatgaa gcggtgcgat accccaaagg agttgaTAAg gcctatggca acctggacag
                              L2 orf start ->
4261 tcttTAAcat ttactaatgc tgcttttgct actaacatac taacataccc tagcatttta
            <- E2 end
4321 tatttttttt tacattttgt atttgctATG gcgcgtgcaa aaacggtcaa gcgagactct
                              L2 cds ->
4381 gtaactcata tttaccaaac ctgcaaacag gcaggcactt gccccccctga tgttattAAT
                                                                 signal ->
4441 AAAgtggaac aaacaacagt tgctgacaat attttaaaat atggcagtgc tggtgtattt
4501 tttggtggcc ttggtattag tacaggccga ggaactgggg gtgctacagg gtacgtgcca
4561 cttggggaag gtcctggtgt ccgtgtcgga ggaACCCCCA CGGTTgtaag gccttccttg
                                        -> E2 bind
4621 gttcctgaaa caatcgggcc cgttgatatt ttgcccattg atacagttaa ccccgtggaa
4681 cctacagcat catccgtggt ccctctaact gagtccacag gcgctgattt acttccaggt
4741 gaagtagaaa caattgctga aatccatcct gtacctgagg ggccatcagt ggatacccct
4801 gtagttacca ctagcacagg ttccagtgct gttttagagg ttgccccaga gcctattcct
4861 ccaacacggg tcaggtttc acgcacacag tatcacaatc catctttca aataataact
4921 gagtctactc cagcacaagg ggaatcgtct cttgcagatc acgttttggt gacatcgggt
4981 tctgggggc aacgaatagg gggtgatata actgacataa ttgagttaga ggaaattcct
5041 agtaggtata catttgaaat tgaagaacca actcctccac gccgcagcag tactccattg
5101 ccacgcaatc aatctgtagg ccgtaggagg ggtttctctt tgactaatag acgtttagta
5161 cagcaggtac aagtggacaa tccattgttt ctaactcaAC CATCTAAGTT agttcgtttt
                                                -> E2 bind
5221 gcatttgata atcctgtttt tgaggaagaa gtgactaata tatttgaaaa tgatctggat
5281 gtctttgaag aacctccaga cagagatttt cttgatgtta gggaattggg acgtccacaa
5341 tattctacaa caccagcggg atatgttaga gtaagcaggt tggggactcg agccactat
5401 cgcactcgct ctggtgcaca gatagggtcg caagtccatt tttacagaga tcttagctct
5461 attaatactg aagatcctat tgaattacaa ttattaggcc aacattcagg tgatgctact
5521 atagtccacg gacctgttga aagccatatt atagatatgg atatttctga aaatccatta
5581 tctgaaagca ttgaagcata ttcacatgat ttattattag atgaaacggt ggaagatttc
5641 agtgggtctc agctggttat aggtaatcga aggagcacaa actcttacac tgttcctagg
5701 tttgaaacta caagaaatgg ttcatactat acacaagaca caaagggata ttatgttgca
5761 tatccagagt cacgtaataa tgcagaaatc atttatccta cacctgatat tcctgtagtc
5821 attatacacc ctcatgacag tacaggggac tttttatttac atcccagtct tcacaggcgc
5881 aaacgtaaaa gaaaatattt gTGAtttgca ttcgagATGg cagtgtggca ctcggctaat
                L1 orf start ->     L1 cds ->
                                  <- L2 end
5941 ggtaaagtat atcttccacc atcgacaccg gtggccagag tccaaagcac cgatgaatac
6001 attcaaagaa caaatatcta ctatcatgca tttagtgaca gattgttaac tgtaggtcat
6061 ccttatttca atgtatacaa tattaatggt gataagcttg aggttcctaa ggtttcagga
6121 aatcaacaca gagtatttcg cctaaaatta ccagatccta acagatttgc attacctgat
6181 atgtctgttt acaaccctga caaagaacgt ttggtttggg cctgtagagg cttagaaata
6241 ggtaggggcc agccattagg tgtacggagt actggtcacc cttatttcAA TAAAgtaaaa
                                                          signal ->
6301 gatacagaaa acagtaatgc atacataaca ttttctaaag atgacagaca ggatacatct
6361 tttgatccta aacagatcca aatgtttat gtaggatgca caccttgcat aggagagcat
6421 tgggataaag ctgttccatg tgcagaaaat gatcagcaaa ctggcctttg tcctcctatt
6481 gaactaaaaa acacatatat acaagatggt gatatggcag acataggttt tgggaacatg
6541 aattttaagg cacttcaaga tagtagatca gatgtcagtt tagacatcgt caatgaaact
6601 tgcaagtatc cagatttttt aaagatgcaa aacgatattt atggcgatgc gtgctttttt
6661 tatgctcgta gggagcaatg ttatgccaga cacttttttg ttagagggg aaaaactggt
6721 gatgacattc cacgtgcaca aattgacaat ggtacataca aaaatcagtt ttacattcca
6781 ggggctgatg gccaagctca aagactata ggaaattcca tgtatttccc aactgttagt
6841 ggctcattag tatccagtga tgctcaattg tttaacaggc ccttctggcc ccaaagagcc
6901 caaggtcata ataatgcat cctgtgggct aatcaaatgt ttatcacagt ggttgacaac
6961 acaagaaata ctaatttcag tatttctgta tataatcagg ctggagcact aaaaagatgtt
7021 gcagactata atgcagatca atttagagaa tatcaaagac atgtagaaga atatgaaata
```

```
7081 tctttaattc tacaactctg taaggttcct ttaaaggcac aggtattggc acagatcaat
7141 gcaatgaact cttcgttatt ggaggattgg cagttaggat ttgttcccac tcctgataat
7201 ccaattcagg acacctacag atatattgac tctttggcta cacggtgtcc agataagaat
7261 cctccgaaag aaaaggaaga cccttataag ggcttacatt tttgggatgt agatttaact
7321 gaaagattgt cattagattt agatcaatat tccttaggca gaaaattttt attccaagct
7381 gggttacaac aaacgACCGT TAACGGTaca aaagcagtgt cttataaagg gtctaataga
                       -> E2 bind
7441 ggaacaaaac gcaaacgtaa aaatTGAggt ctgaccgaaa gtggtacatt tttataaact
                                      <- L1 end
7501 tttacacagt attcaaggaa tgtttgttta ctctgactaa gtataagtct tccaaggata
7561 ccgACCGCAC CCGGTacact cagtcaagtt gttgccaata tagaatcaga tcagtgccaa
        ->  E2-bind
7621 acacaccgtc ttggactcag aacagaccgt gttcgttata acatgctcgg attagggacc
7681 tccccaaaga agatttaatc taCAATCGCT TTTGGCAATC GCATTTGGCA ctgctaaaag
                              ->     overlapping repeat    <-
7741 ACCGTT
     -> E2-bind
```

```
LOCUS       HPV5b           7779 bp ds-DNA   circular  VRL       07-AUG-1991
DEFINITION  Human papillomavirus type 5b (HPV-5b), complete genome.
ACCESSION   D90252
KEYWORDS    complete genome.
SOURCE      Human papillomavirus type 5b DNA from benign lesions of an EV
            patient.
REFERENCE   1  (bases 1 to 7779)
  AUTHORS   Yabe,Y., Sakai,A., Hitsumoto,T., Kato,H. and Ogura,H.
  TITLE     A subtype of human papillomavirus 5 (HPV-5b) and its subgenomic
            segment amplified in a carcinoma: Nucleotide sequences and genomic
            organizations
  JOURNAL   Virology 183, 793-798 (1991)
COMMENT     These data kindly submitted in computer readable form by: Yoshiro
            Yabe Department of Virology, Cancer Institute Okayama University
            Medical School 2-5-1 Shikata-cho Okayama 700 Japan Phone:
            0862-23-7151 x2630 or 2632 Fax:    0862-22-2846

            Yabe et al. studied characterized HPV-5 in lesions of differing
            severity.  They cloned and sequenced HPV-5b from benign lesions of
            a patient with EV.  40% of the genome was amplified in carcinomas,
            and was present in an episomal state.  In the metastatic tumor, only
            the 40% subgenomic region was present, and integrated into the host
            genome.  The segment was determined to be the entire sequences of
            E6, E7, and the noncoding region and portions of E1 and L1, with no
            mutations present. In addition, amplifications of the LCR have been
            reported in HPV-5 associated carcinomas (Deau et al. Virology 184:
            492-503) HPV-5 is associated with macular lesions which frequently
            progress to malignancy. HPV-5 and HPV-8 have also been found in
            significant numbers in squamous cell carcinomas of renal allograft
            patients.  Barr et al. (Lancet 1: 124-9) detected either HPV-5 or
            HPV-8 in nearly 60% of the cases surveyed in the Scotland area.
            HPV-5 is considered to be part of the a$_1$ cluster based on
            phylogenetic analysis.  This cluster includes HPV-5, HPV-8, HPV-47,
            and HPV-12.  Patients with EV tend to have depressed cell mediated
            immunity.  In roughly one-third of EV-associated HPV infection, the
            sun-exposed flat wart-like or macular lesions transform into
            malignant squamous cell carcinomas.  Benign wart scrapings tend to
            be multiply-infected, with as many as six different viral types.
            However, in contrast, EV carcinomas tend to harbor only a few types,
            specifically HPV-5 and HPV-8, and less frequently HPV-14, HPV-17,
            HPV-20 and HPV-47.  These types are rarely detected in lesions
            afflicting the general population.   A key to host restriction of
            these viruses may be in part due to the unusual organization of the
            LCR in these viruses. The LCR of these viruses is short compared to
            the viruses in other groups and contains two EV-specific regulatory
            regions: M33 and M29, both shown to be involved in protein binding.

BASE COUNT     2378 a   1542 c   1752 g   2107 t
ORIGIN      207 bp upstream from beginning of E6 cds
        1 AACGGTaagt agcaagTTCC TTGTTCCTTG tACCAGGTGC GGTattggga ttttgcaatt
         E2 bind <-          -> repeat   <-  -> E2 bind
       61 gtaatggttg ttgccaacta ccataggcac attcaagttt ttgcctgtat cgttttcgta
```

```
                                                        E6 orf start ->
 121 tcctgttaac aatatccaat gtatgtatac ataaataaaT ATATATATAT ATAAgtgtct

                                                signal ->
 181 aagattgggt tattctgtaa tcaggcaATG gctgagggag ccgaacacca acagaaactg
                          E6 cds ->
 241 acagaaaaag ataaggcaga attaccttca accattagag acttagctga aaccttaggc
 301 atccctctta ttgattgtat aataccttgc aattttttgtg gtaaattttt aaattatttg
 361 gaagcctgtg aattcgacta caaaaaactt agcctaattt ggaaagatta ttgtgtgttt
 421 gcgtgttgtc gcgtatgctg tggcgccact gcaacatacg aatttaatca attttatgag
 481 cagacagtgt taggaagaga tattgagtta gcttcaggac tctcgatttt tgatattgat
 541 atcaggtgtc aaacttgctt agcatttctt gacattatag aaaagttaga ttgctgtggc
 601 agaggccttc cctttcacaa ggTGAggaac gcctggaagg gaatctgtag gcagtgtaag
                      E7 orf start ->
 661 cattttatc ATGattggTA Aagaggtcac cgtgcaagat attattctgg agctcagtga
         E7 cds ->        <- E6 end
 721 ggtgcagccc gaagtgctac cagttgacct gtttttgtgaa gaggaattac caaacgagca
 781 ggaaacggag gaggagcctg acatcgaaag gatctcttac aaagttatag ctccgtgcgg
 841 ttgcagacac tgtgaggtca agcttcgcat ttttgtccaa gccacagaat ttggtattag
 901 agctttccaa cagctatTGA ccggagatc gcagctcctg tgtcctgact gtcgcggaaa
              E1 orf start ->
 961 ctgcaaacAT GacggatccT AAtcctaaag gtagtacatc taaagaaggg tttggtgatt
         E1 cds ->            <- E7 end
1021 ggtgtttatt ggaagctgac tgtagtgatg tagaaaatga tttgggacaa ttgtttgaga
1081 gagatacaga ctctgatata tcggatttgt tagatgatac tgaactggag cagggcaatt
1141 ccctggaact atttcatcaa caggagtgtg agcagagcga ggagcaatta caaaaaactaa
1201 aacgaaagta tcttagtcca aaagctatcg cacagcttag tccgcgactt gagtcaattt
1261 cattgtcacc tcagcagaag tctaagcgaa ggctctttgc agagcaggac agcggacttg
1321 agctgacttt aaacaatgaa gctgaagatg ttactcctga ggtggaggta ccggctattg
1381 actctcggcc ggatgacgag ggaggttcag gggatgtaga tatacattat acatcattgt
1441 tgcgttctag caacaaaaaa gccacattaa tggctaaatt taaagagtcg tttggagtag
1501 gttttaatga attgacacgg caatttaaaa gccacaaaac ctgctgtaag gactgggttg
1561 tctctgtata tgcagtgcat gatgatttat ttgaaagctc aaagcagctg ttgcaacagc
1621 attgtgacta tatctgggtc cgtgggatag gtgcaatgtc attatACCTA TTGTGTTtta
                                                  E2 bind ->
1681 aggcgggaaa aaatcgcggg acagttcata agttaattac ctcaatgtta aatgtgcatg
1741 aacagcaaat tttgtctgag ccgccaaaat tgagaaatac agcagctgca ttgttctggt
1801 ataaaggttg tatgggatcg ggggcgttta gccatggacc atatcctgat tggattgccc
1861 aacaaactat attaggtcac aaaagtgctg aggcaagtac ttttgatttt tcagcaatgg
1921 tccaatgggc atttgataat catttattag acgaaccaga tatagcatac cagtatgcaa
1981 ggcttgcacc tgaagatgca aatgcagtag cttggcttgc acataacaac caggccaaat
2041 ttgtgagaga atgtgctgct atggtacgat tttataaaaa gggacaaatg agagacatga
2101 gcatatccga atggatatac acgaaaatta tgaagtagaa ggtgaaggg cactggtcag
2161 atatagtaaa gttattagaa taccaaaata taaactttat tgtattccta actgcattaa
2221 aagaattcct acactcagtg ccaaaaaaaa attgtatttt aatatatggt cctccaaatt
2281 ctggaaagtc atcatttgca atgtctttaa taagagtgtt aaagggtagg gtgttgtcat
2341 ttgtaaactc taagagtcag ttttggctgc aacccctttc agaatgcaag atagctctat
2401 tggatgatgt aacagaccct tgttgggtat acatggtat atatttaaga aatggcttgg
2461 atggacatta tgtttcatta gattgtaaat atagagctcc aacgcaaatg aaatttccc
2521 cattattatt aacatctaac atcaatgtgc atggggaagc caattataga tatttacaca
2581 gtagaataag aggatttgaa tttccaaatc cttttcctat gaaagcagat aatacacctc
2641 agttcgaact aactgaccaa agctggaaat ctttttttac aaggctttgg acacaattag
2701 accTGAgtga tcaagaagag gagggcgagg ATGgagaatc tcagcgagcg tttcaatgct
E2 orf start ->                      E2 cds ->
```

```
2761 ctgcaagatc agctaatgaa catttaTGAa gctgcagaac aaacattgca ggcacaaatt
                           <- E1 end
2821 aaacattggc aaaccttgcg aaaagaagct gtattactct actatgctag ggagaaaggt
2881 gttacaaggc ttggatatca acctgtgcct gtaaaggcag tatcagaaac aaaggctaaa
2941 gaagccatag caatggtgct gcagcttgag tcactacaga catctgactt tgctcatgag
3001 ccatggactc tagttgatac cagcacagaa acatttagaa gcgctccaga aggtcacttc
3061 aaaaaaggcc ccgtccctgt agaagttatt tatgacaatg atccagataa tgccaatttg
3121 tatacaatgt ggacttatgt gtattatatg gatgcggatg ataagtggca taaagcaaga
3181 agtgggggtga atcacattgg catttattat ttacaaggaa cttttaaaaa ctattatgta
3241 ctgtttgctg acgatgcaaa acgatatggt acaactggag aatgggaagT AAaagttaat
                                                    E4 orf start ->
                                                    NH2 terminus unknown
3301 aaggatactg tgtttgctcc tgtcaccagc tccacgcctc cagggtcgcc aggaagacaa
3361 gcagacacag acaccaccgc caagacccccc accacctcca caaccgccgt TGActccacg
                                                    E5 orf start ->
                                                    NH2 terminus unknown
3421 tccagacagc tcaccacatc aaaacagcca caacaaaccg aaaccagagg aagaaggtac
3481 ggacggaggc cctccagcaa gtcaaggaga tcgcaaacgc agcaaaggcg atcaaggtcc
3541 cgacACCGGT CCCGGTctcg gtcccggtcg cgctccaagt cccaaaccca caccacttgg
              -> E2 bind
3601 tccaccacca ggtcccggtc cacgtcggtc ggcaagactc gggccccttac aagcagatcg
3661 cgatccaggg gaaggtcccc aagtacctgc agaaggggag gtggaaggtc acccaggcgg
3721 cgatcaaggt caccctccac ctactcctcc tgcaccacac aacggtcaca gcgggcacgg
3781 gccgaaagtc caacaaccag aggggcccga gggtcgagag ggtcacgagg agggagccgt
3841 gggggggagat tgcggcgacg aggaaggtca tcctcctcct cctcccccgc ccacaaacgg
3901 tcacgagggg ggtctgcTAA gctccgtggc gtctctcctg gtgaagtggg agggtcactt
                      <- E5 end
3961 cgatcagtta gttcaaagca tacaggacga cttggaagat tactggaaga agctcgcgac
4021 cccccagTAA tcattgtcaa aggggcggct aacacactaa aatacttccg caacagagct
                <- E4 end
4081 aaaattaaat acacgggact gtttaggtca tttagtacta cctggtcatg ggtggcagga
4141 gatggcactg agcgtctagg caggcccaga atgctcatta gcttttcttc ctatagtcaa
4201 agaagagatt ttgatgaagc agtgcgatac ccaaaggggag ttgataaggc ctatggcaac
4261 cttgacagtc ttTAAcattt actaatgctg cttctgctac taacatacTA Acatacccta
                <- E2 end                         L2 orf start ->
4321 gcgttttata cttttttaca ttttgtattt gctATGGcgc gtgctaaaag ggtcaagcga
                                        L2 cds ->
4381 gactctgtaa ctcatatttta ccaaacctgc aaacaggcag gtacttgccc ccctgatgtt
4441 attAATAAAg tggaacagac aacagttgct gacaatattc tcaaatatgg cagtgctggt
    signal ->
4501 gtattttttg gtggccttgg tattagtaca ggccgaggaa ctggggggtgc tacagggtac
4561 gtgccacttg gggaaggtcc tggtgtccgt gtcggaggaA CCCCCACGGT Tgtaaggcct
                                                    E2 bind ->
4621 tccttggttc ctgaaacggt tgggcccgtt gatattttgc ccattgatac agttaacccc
4681 gtggaaccta cagcatcatc cgtggttcct ttaactgagt ccacaggcgc tgatttactt
4741 ccaggtgaag tagaaacgat tgctgaaatc catcctgtac ctgaagggcc atcggtagat
4801 acccctgtgg ttaccactag cacaggttcc agtgctgttt tagaggttgc ccccgagcct
4861 attcctccaa cacgggtcag agtttcacgc acacattatc acaatccatc attccaaata
4921 ataactgagt ctactccagc acaaggggaa tcgtctcttg cagatacacgt gttggtgaca
4981 tcaggttctg ggggggcagca aatagggggg gatataactg acattattga gttagaggaa
5041 attcctagta ggtatacatt tgaaattgaa gagccaactc ctccacgccg cagcagtact
5101 ccattgccac gcaatcaatc tgtaggccgc aggaggggggt tctccttgac taatagacgt
5161 ttggtacagc aggtacaagt ggataatcca ttgtttctaa ctcaACCATC TAAGTTagtt
                                                    E2 bind ->
5221 cgtttttgcat ttgataatcc tgtttttgag gaagaagtaa ctaatatatt tgaaaatgat
5281 ctggatgtat ttgaagaacc tccagacaga gattttcttg atgttaggga attaagacgt
5341 ccacaatatt ctacaacacc agcgggatat gtcagggtaa gcagattggg gacccgagcc
5401 actattcgca ctcgctcggg tgcacaaata gggtcgcaag tccatttttta cagagatctt
5461 agctctatta atactgagga tcctattgaa ttacaattat taggccaaca ttctggtgat
5521 gctactatag tccagggacc tgttgaaagc acatttatag atatggatat ttctgaaaat
5581 ccattattcag aaaagcattga agcatattca catgatttat tactagatga ggcggtggaa
5641 gattttgttg ggtcccagct tgttataggt aatcgaagga gcacaaactc ttacactgtt
5701 cctaggtttg aaaactacaag aaatggttca tattataccc aagacacaaa gggatattat
5761 gttgcttatc cggagtcacg taataatgca gaaatcatttt atcctacacc tgacattcct
```

```
5821 gtggtcatta tacacactca tgacaataca ggggactttt atttacatcc cagtcttcgc
5881 aggcgcaaac gtaaaagaaa atatttgTGA tttgcattgc agATGgcagt gtggcactcg
                        L1 orf start ->   L1 cds ->
                                      <- L2 end
5941 gctaatggta aagtatacct tccaccatcg acaccggtgg ccagagtcca aagcaccgat
6001 gaatacattc aaagaacaaa tatctactat catgcattta gtgacagatt gttaactgta
6061 ggtcatcctt atttcaatgt atataatatt actggtgata agcttgaggt ccctaaggtt
6121 tcaggaaatc aacacagagt atttcgccta aaattaccag atcccaacag atttgcatta
6181 gctgatatgt ctgtttacaa ccctgacaaa gaacgtttgg tttgggcctg tagaggctta
6241 gaaataggta ggggccagcc attgggtgta gggagcactg gtcaccctta tttcaataaa
6301 gtaaaagata cagaaaacag taatgcatac ataacatttt ctaaagatgg acagaataca
6361 gcattttcta aagatgacag actgaataca tcctttgatc ctaaacaaat ccaaatgttc
6421 attgtaggat gcacaccttg cataggagag cattgggata aagctgtgcc ttgtgcaaaa
6481 aatgaccagc aaactggcct ttgtcctcct attgaattaa aaaatacata tatagaagat
6541 ggtgatatgg cagatatagg ttttggaaat atgaacttta aggcacttca agatagtaga
6601 tcagatgtca gtttggatat tgtcaatgaa acttgcaaat atccagattt tttaaagatg
6661 caaaatgata tctatggcga tgcctgcttt ttttatgctc gtagggagca atgttatgct
6721 agacactttt ttgttagagg gggtaaaact ggtgatgaca ttccaggtgc acaaattgac
6781 aatggtacat acaaaaatca attttacatt ccaggagctg atggccaagc tcaaaagact
6841 atcggaaatg ccatgtattt cccaactgtt agtggctcat tagtttccag tgatgctcaa
6901 ttgtttaaca ggcccttctg gctccaaaga gcccaaggtc ataataatgg catcctgtgg
6961 gctaatcaaa tgtttatcac agtggttgac aacacaagaa atactaattt cagtatttct
7021 gtatataatc aagctggacc actaaaagat gttgcagact ataatgcaga gcaatttaga
7081 gaatatcaaa gacatgtaga agaatatgaa atatctcttaa ttttacaact ttgtaaggtt
7141 cctttaaagg cggaggtatt ggcacagatc aatgcaatga actcctcttt attggaagat
7201 tggcagttag gatttgttcc cactcctgat aatccaattc aggataccta caggtatatt
7261 gactctttgg ctacacggtg tccagataaa aatcctccaa aagaaaagga agacccttat
7321 aaaggcttac attttttggga tgtagattta actgaaagat tgtcattaga tttagatcaa
7381 tattccttag gcaggaaatt tttattccaa gctggtttac aacacacgAC CGTTAACGGT
                                                          E2 bind ->
7441 acaaaagcag tgtcttataa agggtctaat agaggaacaa agcgcaaacg taaaaatTGA
                                                                     <-
                                                                     L1 end
7501 ggcctgACCG AAAGTGGTac atttttataa acttttacac agtattcaag gaatgtttgt
        E2 bind ->
7561 ttactctgac taagtataag tcttccaagg ataccgACCG CACCCGGTac actcagtcag
                                               E2 bind ->
7621 gttggttgcca atatagaatc agatcggtgc caaacacacc gtcttggact cagaacagac
7681 cgtgttcgtt ataacatgct cggattaggg acttcgccaa agaagattta atctaCAATC
                                                                     ->
7741 GCTTTTGGCA ATCACATTTG GCActgctaa aggACCGTT
          repeat                <-  E2 bind ->
```

```
LOCUS       HPV5d          7746 bp ds-DNA          VRL       15-JUN-1989
DEFINITION  Human papillomavirus type 5d (HPV5d), complete genome which contains
            naturally occurring deletions in the late region of the virus.
ACCESSION   M22961 M18452 M18453 M18454
KEYWORDS    complete genome.
SOURCE      Human papillomavirus type 5d DNA.
REFERENCE   1  (sites)
  AUTHORS   Ostrow,R.S., Zachow,K.R. and Faras,A.J.
  TITLE     Molecular cloning and nucleotide sequence analysis of several
            naturally occuring HPV-5 deletion mutant genomes
  JOURNAL   Virology 158, 235-238 (1987)
REFERENCE   2  (bases 1 to 7746)
  AUTHORS   Ostrow,R.S.
  JOURNAL   Unpublished (1988) Univ. of Minnesota, Minneapolis MN 55455
COMMENT     Ostrow et al. conducted a study to characterize HPV-5 detected
            in lesions of differing severity.  They mapped three naturally
            occurring HPV-5 deletion mutants with deletion sizes of 1.33 kb,
            1.9 kb, and 2.3 kb.  All of these deletions were located within the
            late region of the genome. Other than these deletions, the HPV-5d
            sequence is virtually identical to the HPV-5 prototypic sequence.
            The features table in this entry represents the locations of the
            deletions found in the three mutants.  The first form of HPV-5
            characterized contained a deletion of 1.3 kb and is represented
            by the HPV-5d sequence with the naturally occurring deletion A.
            The second form represents the 1.9 kb deletion, it is composed of
            the sequence with naturally occurring deletions B and C .  The third
            and final form characterized was a 2.3 kb deletion, the 5d
            sequence with naturally occurring deletion D. The third form of
            HPV-5d characterized was derived from a metastatic lesion.  For
            other site information, see entry HPV-5.

            NCBI gi: 333086
BASE COUNT     2376 a   1534 c   1749 g   2087 t
ORIGIN      362 bp upstream of HindIII site.
        1 aacggtaagt tgcaatttcc ttgtaccagg tgcggtattg ggatttcaca attataatgg
       61 ttgttgccaa ctaccatagg catattcaag ttttttgcctg tatcgttttc gtatcctgta
      121 ataatatcca atatatgtat acataaataa atatatatat atataagtgt ctaagattgg
      181 gttcttctgt aatcaggcaa tggctgaggg agccgaacac caacagaaac tgacagaaaa
      241 agataaggca gaattacctt taagtattag agacttagct gaagccttag gcatccctgt
      301 gattgattgt ttaatacctt gcaatttctg tggcaacttt ctaaattatt tggaagcttg
      361 tgaattcgac tacaaaaggc ttagtctaat ttggaaagat tattgtgtgt ttgcgtgctg
      421 tcgcgtatgc tgtggcgcca ctgcaactta tgaatttaac caattttatg agcagacagt
      481 gttaggaaga gatattgaat tagcttcagg actttcaata tttgatattg atatcaggtg
      541 tcaaacttgc ttagcatttc ttgacattat agaaaagtta gattgctgtg gcagaggcct
      601 tccctttcat aaggtgagga acgcctggaa gggaatctgt aggcagtgta agcattttta
      661 tcatgattgg taaagaggtc accgtgcaag atattattct ggagctcagt gaggtgcagc
      721 ccgaagtgct accagttgac ctgtttttgtg aagaggaatt accaaacgag caggaaacgg
      781 aggaggagcc tgacaacgaa aggatctctt acaaagttat agctccgtgc ggttgcagga
      841 actgtgaggt caagcttcgc attttttgtcc acgccacaga atttggtatt agagctttcc
      901 aacagctact gaccggagat ctgcagctcc tgtgccctga ctgtcgcgga aactgcaaac
      961 atgacggatc ctaattctaa aggtagtaca tctaaagaag ggtttggtga ttggtgttta
     1021 ttggaagctg actgtagtga tgtagaaaat gatttgggac aattatttga gagagataca
     1081 gactctgata tatcggattt gttagatgat actgaactgg agcagggcaa ttccctggaa
     1141 ctatttcatc aacaggagtg tgagcagagc gaggagcaat tgcaaaaact aaaacgaaag
     1201 tatcttagtc caaaagctgt cgcacagctt agtccgcgac ttgagtcaat ttcattgtca
     1261 ccccagcaga agtctaagcg aaggctcttt gcagagcagg acagcggact cgagctgact
     1321 ttaaacaatg aagctgaaga tgttactcct gaggtggagg taccggctat tgactctcgg
     1381 ccggatgacg agggaggttc aggggacgta gatatacatt acactgcatt gttgcgttct
     1441 agcaacaaaa aagctacatt aatggctaag tttaaagagt cgtttggagt aggttttaat
     1501 gaattgacac ggcaattcaa aagccacaaa acctgctgta aggactgggt tgtctctgta
     1561 tatgcagtgc atgatgatct atttgaaagc tcaaagcagc cattgcaaca gcattgtgac
     1621 tatatctgga tccgtgggat aggtgcaatg tcattatacc tattgtgttt taaggcggga
     1681 aaaaatcgcg ggacagttca taagttaatt acctcaatgt taaatgtgca tgaacagcaa
     1741 atattgtctg agccgccaaa attgagaaat acagccgctg cattgttctg gtataagggt
```

```
1801 tgtatgggat cggggggcgtt tagccatgga ccatatcctg attggattgc ccaacaaact
1861 atattaggtc acaaaagtgc tgaggcaagt acttttgatt tttcagcaat ggtccaatgg
1921 gcatttgata atcacttatt agacgaagca gatatagcat accagtatgc aaggcttgct
1981 cccgaagacg cgaatgcagt agcttggctt gcacataaca accaggccaa atttgtgaga
2041 gaatgtgcat atatggtacg atttttataag aaagggacaaa tgagagacat gagtatatct
2101 gaatggatat acactaaaat caatgaagta gaaggggaag ggcactggtc agatatagta
2161 aagtttatta gataccaaaa tataaacttt attgtattcc taactgcatt aaaagaattc
2221 ctacactcag tgccaaaaaa aaattgcatt ttaatttatg gtcctccaaa ttctggaaag
2281 tcatcatttg caatgtcatt aataagagtg ttgaagggta gagtgttgtc atttgtaaat
2341 tctaaaagtc agttttggct gcaacccctt tcagagtgca agatagctct attggatgat
2401 gtaacagacc cttgttggat atacatggat acatatttaa gaaatggctt ggatggacat
2461 tatgtttcat tagattgtaa atatagagcc ccaacgcaaa tgaaatttcc cccattatta
2521 ttaacatcta acattaatgt gcatgggggaa actaattata gatatttaca cagtagaata
2581 aaaggatttg aatttccaaa tcctttttcct atgaaagcag ataatacacc tcagttcgaa
2641 ctaactgacc aaagctggaa atcttttttt acaaggctttt ggacacaatt agacctgagt
2701 gatcaagaag aggaggggcga ggatggagaa tctcagcgag cgtttcaatg ctctgcaaga
2761 tcagctaatg aacatttatg aagctgcaga acaaacattg caggcacaaa ttaaacattg
2821 gcaaacctta cgaaaagaag ctgtattact ctactatgct agggagaaag gtgttacaag
2881 gcttggatat caacctgtgc ctgtaaaggc agtatcagaa acaaaggcta aagaagccat
2941 agcaatggtg ctgcagcttg agtcactaca gacatctgat tttgctcatg agccatggac
3001 tctagttgat accagcatag aaacatttag aagcgctcca gaaggtcact tcaaaaaagg
3061 ccccgtccct gtagaagtta tttatgcaaa tgatccagat aatgccaatt gtatacaat
3121 gtggacctat gtgtattata tggatgcgga tgataagtgg cataaggcaa gaagtggggt
3181 gaatcacatt ggcatttatt atttacaagg aactttaaaaa aactattatg tactgtttgc
3241 tgacgatgcg aaaagatatg gtacaactgg agaatgggaa gtaaaagtta ataaggaaac
3301 tgtgtttgct cctgtcacca gctccacgcc tccagggtcg ccaggaggac aagcagacac
3361 aaacaccacc cccgcgaccc ccaccacctc cacaaccgcc gttgactcca cgtccagaca
3421 gctcaccaca tcaaaacagc cacaacaaac cgaaaccaga ggaagaaggt acggacggag
3481 gccctccagc aagtcaagga gatcgcaaac gcagcaaagg cgatcaaggt cccgacaccg
3541 gtcccggtct cggtcccggt cgcggtccaa gtcccaaacc cacaccactc ggtccaccac
3601 caggtcccgg tccacgtcgc tcaccaagac tcgggccctt acaagcagat cgcgatccag
3661 aggaaggtcc ccaaccacct gcagaagggg aggtggaagg tcacccaggc ggcgatcaag
3721 gtcaccctcc acctcctcct cctgcaccac acaacggtca cagcgggcac gagccgaaag
3781 ttcaacaacc agaggggccc gagggtcgag agggtcacga ggagggagcc gtggggggag
3841 agggcggcga cgaggaaggt catcctcctc ctcctccccc gcccacaaac ggtcacgagg
3901 ggggtctgct aagctccgtg gcgtctctcc tggtgaagtg ggagggtcac ttcgatcagt
3961 tagttcaaag catacaggac gacttggaag attactggaa gaagctcgcg acccccagt
4021 aatcattgtc aaagggggcgg ctaacacact gaaaaatgtc cgcaacagag ctaaaattaa
4081 atacatggga ctgtttaggt catttagtac tacctggtca tgggtggcag gagatggcac
4141 tgagcgtcta ggcaggccca gaatgctcat tagctttttct tcctatactc aaaggagaga
4201 ttttgatgaa gcggtgcgat accccaaagg agttgataag gcctatggca acctggacag
4261 tctttaacat ttactaatgc tgcttttgct actaacatac taacataccc tagcatttta
4321 tattttttt tacattttgt atttgctatg gcgcgtgcaa aaagggtcaa gcgagactct
4381 gtaactcata tttaccaaac ctgcaaacag gcaggcactt gccccccctga tgttattaat
4441 aaagtggaac aaacaacagt tgctgacaat attttaaaat atggcagtgc tggtgtattt
4501 tttggtggcc ttggtattag tacaggccga ggaactgggg gtgctacagg gtacgtgcca
4561 cttgggggaag gtcctggtgt ccgtgtcgga ggaacccccca cggttgtaag gccttccttg
4621 gttcctgaaa caatcgggcc cgttgatatt ttgcccattg atacagttaa ccccgtggaa
                                                    deletion B |->
4681 cctacagcat catccgtggt ccctctaact gagtccacag gcgctgattt acttccaggt
4741 gaagtagaaa caattgctga aatccatcct gtacctgagg ggccatcagt ggatacccct
4801 gtagttacca ctagcacagg ttccagtgct gtttttagagg ttgcccccaga gcctattcct
4861 ccaacacggg tcaggggtttc acgcacacag tatcacaatc catcttttca aataataact
4921 gagtctactc cagcacaagg ggaatcgtct cttgcagatc acgtttttggt gacatcgggt
4981 tctggggggggc aacgaatagg ggggtgatata actgacataa ttgagttaga ggaaattcct
5041 agtaggtata catttgaaat tgaagaacca actcctccac gccgcagcag tactccattg
5101 ccacgcaatc aatctgtagg ccgtaggagg ggtttctctt tgactaatag acgtttagta
5161 cagcaggtac aagtggacaa tccattgttt ctaactcaac catctaagtt agttcgtttt
5221 gcatttgata atcctgtttt tgaggaagaa gtgactaata tatttgaaaa tgatctggat
5281 gtctttgaag aacctccaga cagagatttt cttgatgtta gggaattggg acgtccacaa
                                                    deletion D |->
5341 tattctacaa caccagcggg atatgttaga gtaagcaggt tggggactcg agccactatt
5401 cgcactcgct ctggtgcaca gatagggtcg caagtccatt tttacagaga tcttagctct
5461 attaatactg aagatcctat tgaattacaa ttattaggcc aacattcagg tgatgctact
```

```
5521 atagtccagg gacctgttga aagcacattt atagatatgg atatttctga aaatccatta
5581 tctgaaagca ttgaagcata ttcacatgat ttattattag atgaaacggt ggaagatttc
5641 agtgggtctc agctggttat aggtaatcga aggagcacaa actcttacac tgttcctagg
5701 tttgaaacta caagaaatgg ttcatactat acacaagaca caaagggata ttatgttgca
5761 tatccagagt cacgtaataa tgcagaaatc atttatccta cacctgatat tcctgtagtc
5821 attatacacc ctcatgacag tacaggggac ttttatttac atcccagtct tcacaggcgc
5881 aaacgtaaaa gaaaatattt gtgatttgca ttcgagatgg cagtgtggca ctcggctaat
5941 ggtaaagtat atcttccacc atcgacaccg gtggccagag tccaaagcac cgatgaatac
6001 attcaaagaa caaatatcta ctatcatgca tttagtgaca gattgttaac tgtaggtcat
6061 ccttatttca atgtatacaa tattaatggt gataagcttg aggttcctaa ggtttcagga
6121 aatcaacaca gagtatttcg cctaaaatta ccagatccta acagatttgc attagctgat
6181 atgtctgttt acaaccctga caaagaacgt ttggtttggg cctgtagagg cttagaaata
                                              deletion A |->
                                              deletion B  <-|
6241 ggtaggggcc agccattagg tgtagggagt actggtcacc cttatttcaa taaagtaaaa
6301 gatacagaaa acagtaatgc atacataaca ttttctaaag atgacagaca ggatacatct
6361 tttgatccta aacagatcca aatgtttatt gtaggatgca caccttgcat aggagagcat
6421 tgggataaag ctgttccatg tgcagaaaat gatcagcaaa ctggcctttg tcctcctatt
6481 gaactaaaaa acacatatat acaagatggt gatatggcag acataggttt tgggaacatg
6541 aattttaagg cacttcaaga tagtagatca gatgtcagtt tagacatcgt caatgaaact
6601 tgcaagtatc cagatttttt aaagatgcaa aacgatattt atggcgatgc gtgctttttt
6661 tatgctcgta gggagcaatg ttatgccaga cacttttttg ttagagggg aaaaactggt
6721 gatgacattc caggtgcaca aattgacaat ggtacataca aaaatcagtt ttacattcca
6781 ggggctgatg gccaagctca aaagactata ggaaattcca tgtatttccc aactgttagt
6841 ggctcattag tatccagtga tgctcaattg tttaacaggc ccttctggct ccaaagagcc
6901 caaggtcata ataatggcat cctgtgggct aatcaaatgt ttatcacagt ggttgacaac
                                              deletion C |->
6961 acaagaaata ctaatttcag tatttctgta tataatcagg ctggagcact aaaagatgtt
7021 gcagactata atgcagatca atttagagaa tatcaaagac atgtagaaga atatgaaata
7081 tctttaattc tacaactctg taaggttcct ttaaaggcag aggtattggc acagatcaat
7141 gcaatgaact cttcgttatt ggaggattgg cagttaggat ttgttcccac tcctgataat
7201 ccaattcagg acacctacag atatattgac tctttggcta cacggtgtcc agataagaat
7261 cctccgaaag aaaaggaaga cccttataag ggcttacatt tttgggatgt agatttaact
                                              <-| deletion C
7321 gaaagattgt cattagattt agatcaatat tccttaggca gaaaattttt attccaagct
7381 gggttacaac aaacgaccgt taacggtaca aaagcagtgt cttataaagg gtctaataga
7441 ggaacaaaac gcaaacgtaa aaattggagg ctgaccgaaa gtggtacatt tttataaact
7501 tttacacagt attcaaggaa tgtttgttta ctctgactaa gtataagtct tccaaggata
7561 ccgaccgcac ccggtacact cagtcaagtt gttgccaata tagaatcaga tcagtgccaa
          <-| deletion A            <-| deletion D
7621 acacaccgtc ttggactcag aacagaccgt gttcgttata acatgctcgg attagggacg
7681 tcgccaaaga agatttaatc tacaatcgct tttggcaatc gcatttggca ctgctaaaag
7741 accgtt
//
```

```
LOCUS       HPV8           7654 bp ds-DNA   circular  VRL       16-FEB-1987
DEFINITION  Human papillomavirus type 8 (HPV8), complete genome.
ACCESSION   M12737
KEYWORDS    complete genome; major structural protein; minor capsid component.
SOURCE      Human papillomavirus type 8 DNA.
REFERENCE   1  (bases 1 to 7654)
  AUTHORS   Fuchs,P.G., Iftner,T., Weninger,J. and Pfister,H.
  TITLE     Epidermodysplasia verruciformis-associated human papillomavirus 8:
            Genomic sequence and comparative analysis
  JOURNAL   J. Virol. 58, 626-634 (1986)
COMMENT     Draft entry and computer-readable copy of the sequence in [1] were
            kindly provided by H.Pfister, 04-AUG-1986.

            HPV-8 is associated with macular lesions which frequently progress
            to malignancy.  HPV-8 has also been found in significant numbers in
            squamous cell carcinomas of renal allograft patients.  Barr et al.
            (Lancet 1: 124-9) detected either HPV-5 or HPV-8 in nearly 60% of
            the cases surveyed in the Scotland area.  HPV-8 is considered
            to be part of the a$_1$ cluster based on phylogenetic
            analysis.  This cluster includes HPV-5, HPV-12 and HPV-47, in
            addition to HPV-8.  Patients with EV tend to have depressed cell
            mediated immunity.  In roughly one-third of EV-associated HPV
            infection, the sun-exposed flat wart-like or macular lesions
            transform into malignant squamous cell carcinomas.  Benign wart
            scrapings tend to be multiply-infected, with as many as six
            different viral types.  However, in contrast, EV carcinomas tend to
            harbor only a few types, specifically HPV-5 and HPV-8, and less
            frequently HPV-14, HPV-17, HPV-20 and HPV-47.  These types are
            rarely detected in lesions afflicting the general population.  A
            key to host restriction of these viruses may be in part due to the
            unusual organization of the LCR in these viruses. The LCR of these
            viruses is short compared to the viruses in other groups and
            contains two EV-specific regulatory regions: M33 and M29, both
            shown to be involved in protein binding.

            Fuchs et al. note that HPV-8 is unique among the papillomaviruses
            in several respects.  These characteristics include: the small
            noncoding region with little potential to form complex secondary
            structures, a cluster of promoter elements in the 3' half of the
            E1 ORF, and the homology between the E4 ORF and the Epstein-Barr
            virus nuclear antigen 2 protein.

            Stubenrauch et al. ( J. Virol. 66: 3485-93) have deduced the
            transcript structure of HPV8.  They identified a late promoter at
            P$_7535$ which gives rise to mRNAs consisiting of three exons: an
            LCR leader, a short segment from the early region, and the L1 gene.
            There is no classical TATA box recognizable in the late promoter,
            as is the case with the BPV1 late promoter.  Both the HPV8 and BPV1
            late promoters show some sequence similarity to the SV40 major late
            promoter.  They have also identified another promoter at P$_175$.
            They have deduced that this promoter is most likely responsible
            for the early genes.  All splice sites annotated in the text have
            been experimentally determined.

BASE COUNT     2313 a   1551 c   1737 g   2053 t
ORIGIN      HpaI site; 195 bp upstream from beginning of E6 cds
        1 aacgGTaagt ttcatcagtg taccaggtgc ggtatgaaaa tttcttaatc ataagttgtt
      5' sj /\
       61 attgccaaca accatcgtct atagcatgtt tttgcctgta tcgtttttcga tcacaccata
      121 ttgTATATTA aaTAAATAAA taaaTATATA TATATATtgt tacaatgctg tgacttgtgc
            E6 orf start ->        -> signal                     |-> mRNA start
      signal ->   signal ->                                      site from P(175)
                                                                 promoter
      181 aattttccta agcaaATGga cgggcaggac aaggcttcat atttagacac taataaggac
                E6 cds ->
      241 gagctaccct ctactattaa agagttagct gcggctttag gtattccatt gcaggactgt
```

```
 301 tcagtaccgt gcaacttttg tggtaacttt ttggatttct tagaactgtg tgagtttgac
 361 aaaaagagac tgtgcctaat ttggaaaaat tacgttgtta ctgcgtgttg tcgttgttgt
 421 tgtgtagcaa ccgcaacgtt tgaatttaat gaatattatg agcaaactgt gctaggcaga
 481 gatattgaat tagctacagg acgttcaatt tttgagatag acgttaggtc tcaaaactgc
 541 ttgtcatttt tggatatcat agagaaatTA Gattgctgtg ggagaggccg tccctttcat
                        E7 orf start ->
 601 aaagttagag gaggctggaa aggagtttgc aggctttgta agcatttgta tcATGgattgg
                                                           E7 cds ->
 661 TAAagaggtc actgtgcaag attttgtgtt gaagttaagt gagatacaac ctgaagtgtt
           <- E6 end
 721 accagttgac ctgctttgtg aagaggaatt accaaacgaa caggaaacgg aggaggagcT
 781 AGacatcgaa agaactgtat tcaaaattgt tgcaccgtgt ggctgcagct gctgtcaggt
E1 orf start->
 841 caagctacgt cttttttgtca acgcaactga ttcgggtatc aggacctttc aagaattgct
 901 gttcagagac ctacagcttc tgtgtcctga gtgccgcggt aactgcaaac ATGgcggatc
                                                              E1 cds ->
 961 aTAAaggtag tacatctaaa gaagggttaa gtgagtggtg tattttggaa gctgaatgta
           <- E7 end
1021 gtgatgtaga caatgatttt gaacaattat ttgagcgaga tacagactca gatatttcgg
1081 acttattaga taattgtgac ctggatcagg gaaattctct ggaactattt catcaacagg
1141 agtgtgagca gagcgaggag caattacaaa aactaaaacg aaagtatctt agtcctaaag
1201 ctgtcgcgca gctcagtccg cggctccagt caatatcact gtcacctcag cagaaatcca
1261 agcgaaggct ctttgcagag caggacagcg gagtcgagct aactcttaac aatgaagctg
1321 aagatgtttc tcatgaggtg gaggtaccgg ctatagactc tcggccggaa gatgagggag
1381 gatcaggggc tttagatatt gactatacag cattgttgcg gtccagcaac acaaaggcca
1441 cattaatggc aaaatttaaa gaggcatttg gggatggctt taatgaacta acacgccaat
1501 ttaaaagtta caaaacttgc tgtaactatt gggttgtagc tgcttatgca gtgcatgatg
1561 tatatgaaag ctcaaagcag ttattgcaac agcattgtga ttatatttgg gttagaagta
1621 tagctgcaat tacattatat ctattgtctt ttaaggcggg aaaaaatcgc ggtactgtgc
1681 ataaactaat gacctcaatg ctaaatgtgc aagagcagca gatattgtct gagcctccta
1741 aattgagaca cactgctgct gctttgtttt ggtataaagg aggaatgggg acaggaacat
1801 tcacgtatgg ttcataccct gattggattg cacatcaaac aattcttggc catcaaagtg
1861 ctgaagcaag caccctttgat ttttctgtaa tggtacaatg ggcatttgat aataatcatt
1921 ttgaggaggc cgacattgct tatggatatg caaaactagc cccagaagat gctaatgctg
1981 tagcctggct tgctcataat agccaagcta aatttgtaag agagtgtgca gcaatggttc
2041 gattttacaa aagggggcaa atgagagaaa tgaccatgtc tgagtggata tatacaagga
2101 tcaatgaggt tgaaggagga gggcattggt cttccatagt aaaatttgta aggtaccagg
2161 gtataaattt tattacattt ttagctgcct taaaagattt tttgcattcc gtacctaagc
2221 gaaactgttt gttaatctat ggtccaccta atactgggaa atctacattt gctatgtcat
2281 taatacaggt gctgaagggt agagtattat catatgtaaa ttccaaaagc caattttggt
2341 tacaacctct tggtgactgt aaaatagcac tgttagatga tgtgacagac ccctgctggc
2401 tgtacatgga tacatttctg cgaaatggat tagatgggca tgttgtgtca ttagattgta
2461 aatataaagc acccatgcaa attaaatttc ctcccctttt gttaacttcc aatattaacc
2521 tgcatgagga ggctaactat agatatttac atagcagagt cagaggattt gaatttccaa
2581 atcctttttcc aatgaaacca gacaatacac ctgaatttga attaactgac caaagctgga
2641 aatcttttttt tgcaaggctt tggacacaat tagagcTGAg tgatcaagaa gacgagggcg
                        E2 orf start ->
2701 aacATGgaga atctcagcga gcgtttcaat gttctgcaag atcagctaat gaacatttaT
 E2 cds ->
2761 GAagctgcag aacaaacact tgaggcacag attgcgcatt ggctgctttt gcgaaaagaa
           <- E1 end
2821 gctgtgttgc tatatttcgc taggcaaaaa ggcatcacca ggattggata tcaacctgtg
2881 ccgccactag cagtgtctga agcaaaagcc aagcaggcta taggaataat gttgcagctg
2941 cagtcactac agaaatctga gtttgcagat gagccttgga cattagtgga cacaagtata
3001 gagacttata agaatgctcc agaaaatcat tttaaaaaag cgccacacc tgtggaggtg
3061 atatatgata aacagcctga caatgccaat gtatacacta tgtggaagca catttactac
3121 actgatgcag acgataagtg gcacaaaacc accagtgggg ttaaccagac tggcatatac
3181 tatatgcaag gcagcttcag acattattat gttgtgtttg ctgatgatgc acgtagatat
3241 agtgctactg gagaatggga agTAAaaatt aataaggaca ctgtgtttgc tcctgttacc
                    E4 orf start ->
                    NH2 terminus unknown
3301 AGctccaccc cccccggatc accaccagga caagcagaca cagacaccgc cgccaagacc
           /\ 3' sj
3361 cccaccacct ccgctgactc cacgtccaga cagcagcggt cccctgcaaa acagccacaa
```

```
3421 caaaccgaaa ccaaaggacg aagGTacggg agacggccgt ccagcagaac aagaccgcaa
                       5' sj /\
3481 aaagagcaga ggcgatcaag gtcgcgacac cgcaccaggt ctcgctcccg gtcgctctcc
3541 cgggttaggg ccgttggctc caccaccgta tccaggtcca ggtcctcgtc gctcACCAAG
                                                              E2 bind ->
3601 GCAGTTcggc cccggtccag atcgcgatcc agaggacggg ctacagccac ctctaggcga
3661 agggcaggtc gagggtcacc caggcgacgg cgatcaacct caagGTcacc ctccaccaac
                                                   5' sj /\
3721 accttcaaac ggtcacaaag gggaggaggg agacggggaa gaggaagggg cagtaggggg
3781 agacgggaac gatcatcctc cacctccccc accccacca aacgttcaag aggggagtct
3841 tctaggttgc gtgggggtctc tccttctgaa gtgggaagat cagttcaatc tgttagtgca
3901 aaacatacag ggcgacttgg aagactactg gacgaagcta tcgaccccccc agTAAtattg
                                                              <- E4 end
3961 gttcgagggg aggcaaacac attaaaatgc tttcgcaaca gagctagagt tagatatadaga
4021 ggactgttca aatactttag caccacgtgg tcatgggtgg ccggcgatag cactgagcgt
4081 ctaggcaggt ccagaatgct cattctgttt acttcagctg gccaaagaaa ggactttgat
4141 gagactgtga aatacccgaa gggagttgat acatcttatg gtaacctgga cagtctaTAA
                                                              <- E2 end
4201 cattactaac gctgccttgc tactaacaca ctaacatatt cccatttgct ttttactata
4261 tttttTAAtt gtatactgct ATGgcgcgtg ctagacgggt caagagagac tctgtcacac
L2 orf start ->        L2 cds ->
4321 acatttatca aacctgcaaa caggcaggta catgcccccc tgatgttatt AATAAAgttg
                                                             signal ->
4381 agcaaacaac agttgctgac aatatttaa aatatggcag tgctggtgta tttttttggag
4441 gccttggtat aggtACCGGC CGTGGTacag ggggtgttac tgggtacacg ccattaagtg
                        -> E2 bind
4501 aggggcctgg tatccgtgtt ggaaatactc ccacagtggt acggccctca cttgttccag
4561 aagcggtagg tcctatggac atactgccaa tcgacactat tgaccctgta gagccttcag
4621 tctcctctgt ggtgccactc actgaatctt caggagccga cctacttcca ggggaagttg
4681 agacaattgc tgaaattcac cctgtacctg aaggccccac cattgattca cctgtggtga
4741 cgacaagcaa agggtcaagt gccatttag aggtggctcc tgagccaacc cctcctaccc
4801 gtgtcagagt ttcacgtaca caatatcaca acccttcttt tcaaatcata actgattcca
4861 cacctacaca aggagaaagt tctctagcag accacatttt agtgacttct ggctctggag
4921 gccagactat aggtagtgac ataacagatg tcatagaact gcaagaattt cctagtaggt
4981 attcatttga gattgatgaa cctacaccac cacggcaaag tagcacacct attgagagac
5041 ctcaagtagt tggtagacgt agaggtattt cttaacaaa tagacgactg atacagcagg
5101 tggctgttga ggatcctttg tttttatcaa aaccttccaa attagtaagg ttcagttttg
5161 ataatcctgt atttgaagag gaggtcacta atatatttga gcaagatgtg gacatggtcg
5221 aggagcctcc agatagagat tttcttgatg ttcggcagct agggcgccca caatactcta
5281 ctacaccagc tggctatgtc agagttagca ggttaggcac acgtggcacc attcgcacac
5341 gttctggtgc acaaattggg tcgcaggtac attttttacag ggatttaagt tcaatcaata
5401 cagaagatcc tatagaactc caactattag gtcaacattc tggggattct acaatagtac
5461 agggtcctgt ggaaagcact tttgtaaatg ttgatatttc tgaaaatcct ttatctgaaa
5521 gtattcaggc attttcagat gatttactac tggatgaaac tgtagaggat tttagtgggt
5581 cccaactggt aataggtaat agaagaagta caacttctta tactgttcct agatttgaaa
5641 caacaagaag tggctcttat tatgtgcaag acactaaagg ttattatgtg gcttatcctg
5701 aaagtcgcaa taatgaggaa ataatttacc ctacacctga tcttccggtg gtaaTAAtcc
                                                         L1 orf start ->
5761 acacacATGa taacagtggt gacttcttct tacatcctag ccttcgcaga cgcaaacgta
     L1 cds ->
5821 aaagaaaata tttgtTGAttt tgcattacag atggcagtgt ggcaatcggc taccggtaag
                        <- L2 end
5881 gtttatctgc ctccttcaac accagtggcc agggtgcaaa gcacggatga atacattcaa
5941 agaactaaca tctactatca cgccaatact gacagactgc tcactgtagg acatccatat
6001 ttcaatgttt acaacaataa tggtgacaca ttacaggttc ccaaagtatc gggaaatcaa
6061 cacagggtct ttcgcttaaa gttaccagat ccaaataggt ttgcactggc agatatgtct
6121 gtatacaatc cagacaaaga aaggttggta tgggcttgca gaggcttaga aatcagtagg
6181 ggacaaccat taggtgttgg agcaccggc catccctatt ttAATAAAgt gaaagacact
                                                         signal ->
6241 gaaaacagca attcatacac cacaacatct acagatgaca gacaaaatac ttcctttgat
6301 cctaagcaaa tacaaatgtt cattgtgggt tgcacaccct gcattggtga gcattgggaa
6361 aaagccattc catgtgcaga ggaccaacag caaggtctgt gcccacccat tgaactaaaa
6421 aatacagtta ttgaagatgg ggacatggca gatattggtt ttggtaatat gaactttaag
6481 actttacaac agaacagatc tgatgtaagt ctcgacatag taaatgagat atgcaaatat
```

```
6541 cctgatttt  tgaaaatgca  gaatgatgtg  tatggtgatg  cctgtttctt  ttatgcacga
6601 agggaacagt  gctatgctag  acattttttt  gtgagagggg  gtaagacagg  tgatgacata
6661 cctgctgctc  aaattgatga  tggcatgatg  aaaaatcaat  attacattcc  tggtggacaa
6721 gatcaatcac  aaaaggatat  aggtaatgct  atgtatttcc  caaccgtcag  tggctcactt
6781 gtttcaagtg  atgctcaatt  gtttaacagg  cctttctggc  tgcagcgtgc  ccagggtcat
6841 aataatggca  ttctctgggc  taatcaaatg  tttgtcactg  tggtagacaa  cacgcgaaac
6901 accaatttta  gtatttcagt  ttacactgaa  aatggggaac  ttaagaacat  cacagactat
6961 aaatcaaccc  agttcagaga  atatctgaga  catgtagaag  aatatgaaat  ttccctcata
7021 ttacaattgt  gtaagatacc  actaaaggct  gatgttttag  cacaaatcaa  tgccatgaat
7081 tcatcactac  ttgaggaatg  gcaactggga  tttgtaccta  cccctgatac  tccaattcat
7141 gacacctaca  gatatattga  ttctcttgcc  acacgttgtc  ctgataaaag  tcccccctaag
7201 gaaaagcctg  atccatatgc  aaagtttaac  ttctggaatg  tggaccttac  agaacgactt
7261 tccctggatt  tggatcaata  ttcattaggc  aggaagttct  tgtttcaagc  aggtttgcaa
7321 cagacgACCG TAAATGGTac  aaaatctata  tctaggggct  ccgtcagggg  cacaaaacga
               -> E2 bind
7381 aaacggaaaa atTAGattgt  accgttttcg  gtacaaaacc  ataaactttt  acacagtatt
                   <- L1 end
7441 caaggaatgt  ttgtttattc  tgactcagca  tcactctacc  taagaaaccg  accgcacccg
7501 gtacataaag  gtgagtagtt  gccaaaacag  actcagttta  gtgccagaat  agaccatgtt
                                     |-> mRNA start site from
                                         P(7535) promoter
7561 cgttcaaaca  tgctcggatt  aggtcgcctg  ccaaggaagt  attgatcttg  ccaatctatt
7621 ttggcagcgc  ttttggcatc  tccaacggac  cgtt
```

```
LOCUS       HPV9           7434 bp ss-DNA           VRL         04-OCT-1993
DEFINITION  Human papillomavirus type 9 (HPV-9), complete genome.
ACCESSION   X74464
SOURCE      Human papillomavirus type 9 DNA.
REFERENCE   1  (bases 1 to 7434)
  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 7434)
  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-9 was isolated by Kremsdorf et al. in 1982 (J Virol 43:
            436-47) and subsequently sequenced by Dr. H. Delius.  It has
            been associated with both macular and flat wart-like lesions of EV,
            a multifactorial disease, and has also been identified in a
            keratoacanthoma.  HPV-9 is considered to be part of the b cluster
            based on phylogenetic analysis.  This cluster includes HPV-15,
            HPV-17 and HPV-9.  Patients with EV tend to have depressed cell
            mediated immunity.  In roughly one-third of EV-associated HPV
            infection, the sun-exposed flat wart-like or macular lesions
            transform into malignant squamous cell carcinomas. Benign wart
            scrapings tend to be multiply-infected, with as many as six
            different viral types.  However, in contrast, EV carcinomas tend to
            harbor only a few types, specifically HPV-5 and HPV-8, and less
            frequently HPV-14, HPV-17, HPV-20 and HPV-47.  These types are
            rarely detected in lesions afflicting the general population.   A
            key to host restriction of these viruses may be in part due to the
            unusual organization of the LCR in these viruses.  The LCR of these
            viruses is short compared to the viruses in other groups and
            contains two EV-specific regulatory regions: M33 and M29, both
            shown to be involved in protein binding.


BASE COUNT     2363 a   1393 c   1654 g   2024 t
ORIGIN      220 bp upstream from beginning of E6 cds
        1 ccgcaggcaa ccgccaattt cactgccaag gttcgttggc agaccgtcct ggcttcaaaa
       61 cgACCGATAA CGGTaagtct tggcacgtag gtggttattt gatcgttggg atgattgtgg
            -> E2 bind
      121 ttaacaacaa tctacataca cattttcata tgACCGCCTT CGTTaataag cttatataga
                                                  -> E2 bind
      181 cataaatata TAAggtgccA TGtatttaac agagcagatt atggacaggc caaaacctag
                  E6 cds ->
          E6 orf start ->
      241 aacagtaaag gaactagcag acactcttgt gattccttta atagatttgt tgataccttg
      301 taaattttgc aatagatttt tatcttattt tgagctactt aattttgatc acaagtgttt
      361 acagcttatt tggacagagg aggatttggt gtatggactc tgtagtagct gtgcttatgc
      421 gtctgcacag ttagaattta cacatttttt tcaatttgct gtagttggaa aagatataga
      481 aactgtagaa ggaacagcta ttggaaatat ttgtattagg tgtcgctact gtttttaagtt
      541 attagactta gtggagaagt tagctacatg ctataagttt gagcagtttt aTAAggtcag
                                                             E7 orf start ->
```

```
 601 aaacagctgg aaaggattgt gcagacactg tgggtcggta gaATGAttgg gaaagaagct
                                                      E7 cds ->
                                                               <- E6 end
 661 actatACCAG AGGTGGTtct agaactgcaa gagcttgtcc aacccactgc tgACCTGCAT
          -> E2 bind                                          -> E2 bind
 721 TGTTacgaag aattgacaga agaacctgca gaggaggagc agtgtctcac tccctacaag
 781 atcgtagctg gctgtggttg cggtgcaaga cttcgtttat acgtgcttgc tacaaattta
 841 ggaattcgag cgcaacagga acttttgctg ggtgatatac aactggtgtg tccggagtgc
 901 cgaggcagac ttcgccATGa gTGAcaataa aggtactaaa ttagatccta aagaatgctg
              E1 cds ->      <- E7 end
 961 tagtgcttgg ttatcgttag aagcagaatg ctctgattct agtttagatg gtgatttgga
1021 aaaattattt gacgaaggga cagactctga tatttcagac ctaatagatg atggagatgc
1081 tgtacaggga aactcccgcg aactgttttg ccagcaagag agtgaggaaa gcgagcaaca
1141 aacacaattg ctaaaacgaa agtatatcag tccccaagct gtttttgcagc ttagccctca
1201 actggagtct atctctttgt cgcctcagca taaacctaaa aggagattat ttgaacaaga
1261 cagcggacta gaatgttctg taaatgaagc tgaagatctt tctgaaacac aggtggaaga
1321 ggtaccggcc aatccaccaa caacagctca gggaactaag ggcttgggaa ttgttaaaga
1381 tttacttaaa catagcaatg tgaaagctgt attaatggct aagtttaaag aggcgtttgg
1441 tgtggggttt gctgagctaa caagacaata taaaagtaac aaaacatgct gtagagattg
1501 ggtaattgct gtgtatgctg tgaatgatga cttaattgaa agctctaaac aattgttatt
1561 gcagcattgt gcttatattt ggctacatta tatgccacca atgtgtttat atttattatg
1621 ttttaacgta ggcaaaagta gagaaactgt atgtagacta ttaagcactt tgctgcaagt
1681 atctgaagtg caattattaa gtgagcctcc aaagttgcga agtgtgtgtg ctgcattatt
1741 ttggtataaa ggaagtatga accctaatgt atacgcacat ggtgcgtatc ctgaatggat
1801 acttacacaa acactaatta atcaccaatc tgcaaatgct acacaatttg acttatcgac
1861 aatgatacaa tttgcctatg atcatgaata ttttgatgaa gctaccattg catatcaata
1921 tgcaaagctg gctgaaacag atgctaatgc cagggctttt ttacaaagta acagtcaagc
1981 cagactagta aaagaatgtg caaccatggt gagacattac atgaggggag agatgaaaga
2041 aatgagtatg tccacatgga tacatagaaa actgcttaca gtggaaagca atgggcaatg
2101 gtcagatata gtacggttta ttagatacca ggatattaat tttattgaat ttctaacagt
2161 atttaaagca tttctgcaaa acaaaccaaa gcaaactgt ttattatttc atggaccacc
2221 tgacacggga aaatcaatgt ttacaatgtc actaatatct gtgttaaaag aaaaggtact
2281 gtcatttgcc aattgcaaaa gtactttttg gctacaacct atagctgata ctaaaacttgc
2341 tttaattgat gatgtaacac atgtgtgttg ggaatatata gatcagtact taaggaatgg
2401 attggatggc aattatgtat gtttagatat gaaacataga gcaccttgtc aaatgaaatt
2461 tccACCCCTT ATGTTaacgt ctaacataga tattactaaa gaccaaaagt acaaatattt
          -> E2 bind
2521 gcacagcaga gttaaatcct ttgctttcaa taacaaattt ccacttgatg ctaatcacaa
2581 accacaattt gaacttactg accaaagctg gaaatctttt tttaaaaggc tttggcacaca
2641 gttagatcTG Agtgatcaag aagacgaggg agaggATGga aactctcagc gcacgtttca
     E2 orf start ->                     E2 cds ->
2701 atgcactgca agagacttta atggacctgt aTGAatcagg tcgagaggat ctacaaagtc
                                                 <- E1 end
2761 agattgacca ctggcagact ttaagacaag agcaaatact tttgcattat gccaggaaaa
2821 atggagttat gcgattgggg taccaacctg tacctccgtt ggctaccagt gaacagaaag
2881 ctaaagatgc tattggcatg gttttactat tgcaaagcct tcaaagatca gcttatggac
2941 aggaaccttg gacactggca caaactagtc ttgaggcggt acgcagtcca cctgcatatg
3001 cctttaaaaa gggtccacaa aatattgaag tagtttatga tggagatcct gataatgtta
3061 tgagctatac tatatggaac tttatatatt atcagactgt taatgatact tgggaaaaag
3121 ttcaaggtca cgtggattat tttggagcct attactttga agggactgTA Aaaacatatt
                                                           E4 orf start ->
3181 atattaactt tgacaaagAT Gcagccaggt atggcagaac tggcgtttgg gaagtgcatg
                E4 cds ->
3241 ttaacaagga cattgtgttt gcccctgtta ctagctcttc gccaccaact ggagacgggg
3301 gagagacctc caagcacacc ctttccaggt cggggtcgcc aacaacatcg cgactccctg
3361 ccaccaccgt gcccaccgga ggatccagga catcatcccg acgataccaa cgaaaagcct
3421 ctagccccac caccaggaag aaaagacaga gacaaggaga aggagaagga gaaggagaag
3481 gagaagaaac caactacagg agacaaaggt ccagatccaa gggtcgaaca gaaaccgaaa
3541 ggggagggga gcgacggaga cgaggaaggt cctcctccgc agactccact accccccaccg
3601 acaggcgaag ggaagggga ggtggaaggg ggcccacgac caggtcccag tcccgttccc
3661 gttcccgctc ccactcccgg tcgcggtccc gaggagggac tgcttccagg gttggcgtct
3721 cgcctgatga agtgggaaca cgagttcgat cagttggtgc aggacatcac gggagacttg
3781 cacgattact ggctgaggct aaaagacccc catTAAtgct gttgcgtggc gacgccaatg
                                                      <- E4 end
```

```
3841 tgcttaagtg ctatcgcttt cgggaacgca aaaaaaaaag aggcttagta aaatattata
3901 gtactacgtg gtcatgggta ggggaagaca gttgtgatag agttggaaga gcgcgaatga
3961 ttttagcctt tgacacatat gagcacagac aacaattcat taggactatg aaaattaccac
4021 ctacagtaga ttggtcttta ggaaatgttg atgatctgTA Agctttacta acgctaacgc
                                             <- E2 end
4081 tggcattgct acTAAcccat actaactaac aaacccatac taactaacAT Ggttcgtgca
        L2 orf start ->                        L2 cds ->
4141 aaacgtacta aacgtgcctc tgttacagat atatacagag gctgcaaagc tgctggtaca
4201 tgtccaccag atgtaattAA TAAAgtggag cacacaacta ttgctgataa aattttgcaa
                       -> signal
4261 tatggaagtg ctggtgtgtt tttcgggggc ttgggaataa gtacaggccg tggcactggt
4321 ggtgccactg gctatgttcc attaggggaa gggccaggag tccgtgtagg tggcacccccc
4381 actatagttc gccctggggt gatacctgaa ataattggcc caactgatct aattcctttta
4441 gacacagtca gaccaattga ccccacagca cccagtattg tcacaggcac tgacagcact
4501 gttgaccttt tacctggtga aatagaatca attgctgaga tacacccagt accagtggac
4561 aatgctgtag tagatactcc agttgtaaca gaaggtagaa gaggctcgtc tgccatttta
4621 gaggtggctg acccaagccc tcctatgcga acccgtgttg cacgaactca ataccataat
4681 ccagctttttc aaattatttc tgagtctaca cctatgtcag gtgaatcttc cttagcagat
4741 catattatag tttttgaagg atctgggggc cagctagtag gtggtcctag ggaatcatac
4801 acagcatctt ctgaaaacat agaattacaa gaatttccta gtagatatag ttttgaaata
4861 gatgaaggaa cacctcctcg gactagtaca cctgtccaaa gagcagtaca atcattatct
4921 agtctgcgta gagctctata aacagacgt cttacagaac aagtggctgt gacagatcca
4981 ttattttaa gtaggccttc tcgtttagtt caatttcagt ttgataatcc agcatttgaa
5041 gatgaggtca cacaaatatt tgaaagagat ctaagtactg ttgaggagcc tccagatagg
5101 caattttag atgtacaacg cctttagtagg ccttttatata cagaaacacc tcaggatat
5161 gttcgggtta gtagactagg ccgaagagca acaatccgca cacgtagtgg tgcacaggtg
5221 ggcgcacagg ttcatttcta cagggactta agcaccatta acacagaaga acctatagaa
5281 atgcaattat gggggaaaca ctcaggtgac agtaccatag tacaaggccc agttgaaagt
5341 tcaattgttg atgttaatat tgatgaacct gatggttttgg aggtgggaag acaggaaACC
                                                          E2 bind ->
5401 CCTTCTGTTg aagatgtgga ttttaattct gaagacttac tgttagatga gggtgtagaa
5461 gatttagtg ggtctcagct agtcgttggc acacgccgca gtacaaatac attaacagtg
5521 ccacgctttg aaactccaag ggacactagt ttttatattc aggatataca aggctacaca
5581 gtgtcctatc ccgagtctag acaaaccaca gatataattt ttcacatcc tgacaccccc
5641 acagtagtaa tccacatcaa TGAtacatca ggagattatt atttacaccc aagtctccaa
                L1 orf start ->
5701 aggaaaaaac gcaaacgcaa atatttaTAA ttttgttttt gcagATGtca ttgtggcttc
                              <- L2 end  L1 cds ->
5761 cagcaagtgg taaggtatat ttgccaccag caacaccagt ggcgagagtt caaagcaccg
5821 atgaatatgt ggaaagaaca aatatttttt atcatgcaat tagtgaccgt ttgctaacag
5881 tgggtcatcc atattatgat gtccgctcag gcgacggaca aaggattgaa gtccctaaag
5941 tgtctggtaa tcagtatcgg gcctttagaa ttagcttacc tgatccaaat aggtttgctt
6001 tagcagatat gtcagtttat aatcctgata aggaacgtct agtttgggcc tgtagaggta
```

```
6061 ttgaaatagg cagaggacaA CCTTTAGGGG Ttggaacatc aggtcaccca ttatttaata
                         -> E2 bind
6121 aggttagaga cacagaaaac tctagcaatt atcaaggcac aacaatggat gacaggcaaa
6181 acacatcttt tgaccccaaa caggtacaaa tgttcattat aggatgtatt ccatgcttag
6241 gagaacactg ggataaagcc aaagtgtgtg aaaaggatgc taataatcaa ctaggcttat
6301 gtcctcctat agaattaaga aacacagtaa ttgaggatgg ggacatgttt gatattggat
6361 ttggaaatat caacaataag gaactgtcct ttaataagtc tgatgtaagc ttagatattg
6421 ttgatgaaac ctgcaaatat ccagactttc taacaatggc aaatgatgtt tatggagatg
6481 catgtttctt ttttgcaaga agagaacaat gttatgccag gcattattat gttagaggag
6541 gttcagttgg tgacgctgtt cctgatggtg cagtaaacca ggatcataat ttctttttgc
6601 cagcaaaaag tgatcaacaa caacgaacaa tagctaattc cacctactat cctacagtaa
6661 gtgggtcatt agtaacttca gatgctcaat tgtttaatag gccatttttgg ctccaaagag
6721 cacaaggtca caacaatggc attttatggg gtaatcagat atttgttaca gtggcagaca
6781 atacacgtaa caccaatttt accattagtg tgtctacaga ggcagctcaa acagaagaat
6841 ataatgccaa taatattaga gaatatttaa gacatgttga agaatatcag atttcattaa
6901 tcttacagtt gtgtaaagtg cctttagtag ctgaagtatt atcccagata aatgcaatga
6961 actcaggtat tttagaggat tggcaattag ggtttgttcc aactcctgaa aatgctgttc
7021 atgatatcta cagatatatt gattcaaaag ccacaaaatg cccagatgct gttgagccta
7081 cagaaaaaga agatcccttt gccaaatact cattttggaa agtggatcta actgaaagat
7141 tatcgttgga tcttgatcaa tatcctttag gtagaaaatt tctttttcaa gctggtttgc
7201 aaacacgaaa acgtcctatt aaaacatctg ttaaaacatc taaaaatgct aagagaaggc
7261 gaaccTAACC GATATCGGTt tccaataaaa tttaagttat ccaatttggt atgtgaagca
             <- L1 end
              -> E2 bind
7321 ttttttaacc atcttcgtga ctaaaccgta caagtcaaca cagagcgACC GCACCCGGTt
                                                           -> E2 bind
7381 tatctgatta taaagtgcac ctggtgcaat ttgaacaata ctatcgtgga atca
```

```
LOCUS       HPV12         7673 bp ds-DNA           VRL        04-OCT-1993
DEFINITION  Human papillomavirus type 12 (HPV-12), complete genome.
ACCESSION   X74466
SOURCE      Human papillomavirus type 12 DNA.
REFERENCE   1  (bases 1 to 7673)
  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 7673)
  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-12 was isolated by Kremsdorf et al. in 1983 (J Virol 48:
            340-51) and subsequently sequenced by Dr. H. Delius.  It has
            been associated with both the benign macular and flat wart-like
            lesions of EV, a multifactorial disease.  HPV-12 is considered
            to be part of the a$_1$ cluster based on phylogenetic
            analysis.  This cluster includes HPV-5, HPV-8 and HPV-47, in
            addition to HPV-12.  Patients with EV tend to have depressed cell
            mediated immunity.  In roughly one-third of EV-associated HPV
            infection, the sun-exposed flat wart-like or macular lesions
            transform into malignant squamous cell carcinomas.  Benign wart
            scrapings tend to be multiply-infected, with as many as six
            different viral types.  However, in contrast, EV carcinomas tend to
            harbor only a few types, specifically HPV-5 and HPV-8, and less
            frequently HPV-14, HPV-17, HPV-20 and HPV-47.  These types are
            rarely detected in lesions afflicting the general population.   A
            key to host restriction of these viruses may be in part due to the
            unusual organization of the LCR in these viruses. The LCR of these
            viruses is short compared to the viruses in other groups and
            contains two EV-specific regulatory regions: M33 and M29, both
            shown to be involved in protein binding.

BASE COUNT     2375 a   1531 c   1714 g   2053 t
ORIGIN      199 bp upstream from beginning of E6 cds
        1 ttgtACCAGG TGCGGTacga tttcccaata gcacattata ctagattgtt gttgccaact
                   -> E2 bind
       61 accatcatca gtttcaagtt tttgcctgta tcgttttcgt atcatactaa ttctgtatat
      121 aattaaataa ataaataaat atatatatat atatatataa tgtataaggc ttggttcttt
      181 tgcaatgTGA ttgggacaaA TGgcacagca ggccgatcag cagacagtga cagacagtac
          E6 orf start ->E6 cds ->
      241 gcctgagctg cccacaacta ttaaagagtt agctgacctt ttagatatac ctttagttga
      301 ctgtttggta ccttgcaatt tttgcggaaa gttcttagat tttctggaag tgtgtgattt
      361 tgacaaaaag cagctaacac taatttggaa aggtcatttt gttactgctt gctgtcgaag
      421 ttgttgcgca gctactgcaa tatatgaatt taatgaattt tatcaacaaa cagtgctagg
      481 tagagatata gagcttgcta ctggaaaatc tatatttgac ttaaagataa ggtgtcagac
      541 gtgcttgtca tttttagata caattgaaaa gttagacagc tgtggtcggg gccttccgtt
      601 ccacaaggtt agagacaggt ggaagggaat ttgcagacag tgcaagcatt tgtatcTGAa
                                                             E7 orf start ->
      661 taATGatcgg TAAagaggtc accgtgcaag attttacctt ggagcttagt gagctgcagc
          E7 cds ->         <- E6 end
      721 ctgaagtgtt accagttgac ctgctttgtg aagaggaatt accaaacgag caggaaacgg
      781 aggaggagtc agatatcgac aggactgtat tcaaaatcat tgcaccgtgt ggctgcagct
      841 cctgtgaggt caaccttcgt atttttgtca acgcaactga tactggcatt aggaccctac
      901 aggacctgcT GAtcagtgac ctgcagctgc tgtgcccaga gtgccgtggt aactgcaaac
          E1 orf start ->
      961 ATGgcggatt cTAAaggtag tacctctaaa gaagggttaa gtgattggtg tattttggaa
          E1 cds ->          <- E7 end
     1021 gcagaatgta gtgatttaga gaatgatttt gaacagttat ttgagcaaga tacagactcc
     1081 gatgtatcgg acttgctaga taatggtgaa cttgaacagg ggaattctct ggaactattt
     1141 catcaacaag agtgtgagca aagcgaggag caattacaaa ttctaaaacg aaagtatctt
     1201 agtccaaaag ctgtcgcgca gcttagtccg cgtctcgagt cgatatcgtt gtcacctcag
```

```
1261 cagaaatcga aacgaaggct ctttgcagag caggacagcg gactcgagct atctctaaac
1321 aatgaagctg aagatgtttc tcctgaggtg gaggtaccgg ctatagactc tcggccggta
1381 gatgagggag gatcaggggc catagatatt gattatctgt cattgctgcg tagtagcaat
1441 attaaagcca cgttaatggc aaaattcaaa gagtcatttg gggtaggctt taatgaattg
1501 actcgccagt ttaaaagtta caaaacctgt tgtaacgatt gggtttttagc tgtgtatgca
1561 gttcatgatg atctatttga aagctcaaag cagctgttgc aacagcattg tgactatatc
1621 tgggtccgtg ggataggagc tatgacactt tACCTATTGT GTTtcaaagc gggaaaaaat
                                        -> E2 bind
1681 cgcggtactg tgcataagtt aatgacatca atgctaaatg tgcaagaaca gcagattttg
1741 tctgagcctc ctaagttaag aaatacagct gctgcattgt tctggtacaa aggtggtatg
1801 gggtcaggcg catttaccca tggcacatat cctgattgga ttgcacatca aacaattttg
1861 ggccatcaaa atgctgaagc aagcacattt gattttttcag ccatggtcca atgggccttc
1921 gataataatt acttagaaga accagatata gcttatcaat atgccaagct tgcaccagaa
1981 gatagcaatg cagtagcatg gctagcacat aatcaacaag ctaaatttgt aagagagtgt
2041 gcagcaatgg tacggtttta taaaaaagga caaatgaaag aaatgagtat gtcagagtgg
2101 atacacacaa aaattaatga agtagaaggt gaagggcatt ggtcagatat agtaaaattt
2161 ttaagatatc aagatgtaaa ctttattacc tttttggcag catttaaaaa cttttttgcat
2221 gctgtaccaa aacacaattg tattcttata tatgggcccc ctaactctgg aaagtcatca
2281 tttgcaatgt cactgataaa agtttttaaaa ggtagagtgt tgtcatttgt gaattccaaa
2341 agtcaatttt ggttacaacc tcttggagaa agtaaaattg cattgttgga tgatgttaca
2401 gacccttgct gggtatatat agacacatat ttaagaaatg gcttagacgg ccattttgtt
2461 tctttagatt gtaaatataa agcgcccgta caaattaagt ttcctccatt gttactcaca
2521 tctaatatta atgttcatgg agaaacaaat tatagatatt tacatagtag gataaaggga
2581 tttgaatttc acatcccctt tcctatgaaa ccagacaata cacctcagtt tcagttaact
2641 gaccaaagct ggaaatcttt ttttgaaagg ctttggacac aattagacct gagTGAccaa
                                                            E2 orf start ->
2701 gaagaggagg gccaacATGg agaatctcag cgagcgtttc aatgttctgc aagatcagct
              E2 cds ->
2761 aatgaacata taTGAagctg cagaacatac acttgagaca cagattgccc attggacact
                <- E1 end
2821 tttgcgaaga gaagctgttt tgctttatta tgctaggcaa aagggtatta caaggcttgg
2881 ataccaaccg gtgcctcat tggcagtttc tgaagcaaaa gcaaaagagg ctatagggat
2941 aatgctgcaa ttgcaatctc tacagaagtc tgagtatgct tcggagaact ggacattagt
3001 ggacacaagt gcagaaacgt ataacaatgt tccagaacat cactttaaaa agggtccagt
3061 gcctgtggag gtcatttatg ataaggagcc agaaaatgca aatgtgtata caatgtggaa
3121 gtatgtgtat tatatggacc cagaggatgt atggcataaa ACCACAAGTG GTgTAAatca
                                                        E4 orf start ->
                                                        -> E2 bind
3181 gacgggcatt tactatttac ATGgagactt taaacactat tatgtacttt ttgctgatgg
                E4 cds ->
3241 tgcacgaatg tatagtaaaa ctggacaatg ggaagttaag gttaataagg aaactgtgtt
3301 tgctcctgtc accagctcca caccacccgg gtcaccagga caaagagacc cagacgccac
3361 cagcaagacc cccgccacct cctccgactc cacgaccaga tccagtgaca acagtcaca
3421 acaagccgac cccaggagga aggatacgg acgacgacca tctagcagaa caaggcgaca
3481 ggaaacgcag caaaggcgat caaggtcgcg ataccgctcc cagtctaact cccggtcgcg
3541 ctcccagtcc caaaccaggg cccttggcgc cacctccgta tccaggtcct ccaggtcccc
3601 gtcggtcaca cagattagga accggaggtc gcgatcgcaa tccagaggaa ggggggggtcg
3661 agggtcatcc accgacacca ccactaagcg gcggagatcc agatcatcct cctccaacac
3721 cagaaaacgg tcacaacggg gagaaagagg acggggagaa aggggcggta gggggaagcg
3781 acgagaccga tcatcctcca cctccccccac cccaaaacga tcccgagcag ggtctaggtc
3841 tagcaggcaa cgtggcgtgt ctcctgagca gtgggaaga tctcttcaat ctgttagttc
3901 aaaacataga ggacgacttg ggagattatt ggaagaagct ctcgacccccc cagTGAtcat
                                                                    <- E4 end
3961 ttgcaaaggg ggggcaaaca cgctaaaatg ctttcgcaac agagctagac acaaatatac
4021 agggctgttt aaggctttta gcacaacatg gtcatgggtg gctggagata gcactgagcg
4081 tctaggcagg cccagaatgc tcattagctt tacttcaact aatcagagaa aggattttga
4141 tgagactgtg aaatatccga agggagttga aactgcatat ggcaacttag acagcctaTA
4201 Acataactaa ctttgctttg ctactaacca cacTAAcaat ttttttacat ttatttttac
      <- E2 end                    L2 orf start ->
4261 gtttttatta ttgctATGgc gcgtgctaag cgggtcaagc gagactctgt tactcatata
              L2 cds ->
4321 tatcaaacct gcaaacaagc aggcacttgc cccccctgatg ttctcAATAA Agtggagcaa
                                                        signal ->
4381 acaacagttg ctgacaatat tttgaaatat ggtagtggtg tgtgtcttttt tggaggcctc
```

```
4441 ggtattggtA CCGGTCGCGG Tactggaggt gtcactggat acagacctct acctgaaggg
                -> E2 bind
4501 cccggtatcc gtgttggagg gactcccacg gttgtaaggc cttcacttgt tcctgaatct
4561 gtgggtccag cagatatatt accaatagac actatcgatc ctgtggagcc cactgcttcc
4621 tcggtagttc cattaactga atcctcagca actgatctac ttccaggaga agttgagaca
4681 atagctgaga ttaatcctgt ttcagagggc cctacgattg attcacctgt tgtgacaaca
4741 agcagaggct ccagtgcaat attggaagtt gcaccagacc ctatacctcc aacacgtgtt
4801 agagtggcac gcacacagta ccataatcct gcttttcaaa ttataacaga gtccacacct
4861 gctcaaggtg aaacatcttt ggcagaccac atcctggtca catctggctc aggaggtcaa
4921 actataggta gtgacataac agacattatt gaattacagg aaatacccag tagatactcc
4981 tttgaaatag aagagccaac cccccccgg caaagcagca ctccacttca gaggacacaa
5041 accactggcc gacgtagagg agtgtcccta acaaatagaa ggctagtaca acaagtgcaa
5101 gttgataatc ctttatttat agataaACCT TCTAAGTTAg tacgctttc atttgataac
                       -> E2 bind
5161 cctgtatttg aggaggatat aacaaatatt tttgaacagg acttagaaac atttgaggag
5221 ccacctgata gggatttcct tgatattaaa aagctaagtc gacctcaata ctctacaaca
5281 cctgctggat atgttagggt cagtaggcta ggcacgcggg gtaccattcg cacacgttcc
5341 ggagctcaaa taggatcaca ggttcatttt tatagagact taagttctat tgattcagaa
5401 gaccctattg agctacaatt attaggcaaa cattcaggtg atgcaaccat agtgcaaggc
5461 actgttgaaa gtacatttgt agatatggac atagcggaag atcctttgtc tgaaagtatt
5521 gaggctcatt ctgatgacct attacttgat gaagctgtgg aggattttag cgggtcacaa
5581 ttagttattg gaaacagacg cagcacaaca tcatacactg tgccccgttt tgaaactaca
5641 aggagcagtt cttattatgt gcaagataca caaggttatt atgtggctta tcctgaacat
5701 agaaatactg cagaaatcat ttatccaaca ccagatatac ctgtggtagt tatacataca
5761 catgataata gtggtgactt ttatttacat cctagcctgc ggaggcgcaa gcgTAAaaga
                                                    L1 orf start ->
5821 aaatatttgT GAtttacttg cagATGgctg tgtggcaagc ggcccatggt aaggtctatc
                     L1 cds ->
                <- L2 end
5881 taccaccatc aacaccagtg gccagggtgc aaagcacgga tgaatacatt caaaggacta
5941 acatctacta tcacgccaat actgacagac tactcactgt aggacatcca tatttcaatg
6001 tttatgataa cactggtaaa aaattggagg ttcctaaagt gtcaggaaat caacacagag
6061 tatttcgcct caaattgcca gatccaaaca gatttgcttt agctgacatg tctgtatata
6121 atcctgatag ggaaaggttg gtctgggcct gcagaggatt ggaaataagt aggggtcagc
6181 ctttaggcgt tggaagtact ggacatccct attttaacaa aattaaagac acggaaaact
6241 caaataacta tgccacaggc agtaaggatg atagacagaa cacatcattt gatcctaaac
6301 aaatccaaat gtttatagtg ggctgtacac cttgtgtggg agaacactgg agagaaagcct
6361 taccctgcgg ggatgcacca gctgaaaatg gtgtttgccc tcccatagag ttaaagaaca
6421 ctttcattga agatggtgac atggcagata ttggtttttg gaacatgaat tttaaaacat
6481 tgcaacaaaa cagatctgat gtcagccttg acatagtgaa tgaaacttgc aaatatccag
6541 attttttgaa aatgcaaaac gatgtatatg gagatgcttg ctttttttat gctcgtagag
6601 agcagtgtta tgccagacat ttctttgtga gaggaggtaa aacaggtgat gacattcctg
6661 acgcacaaat tgatgatggc aatatgaaaa atcagtatta cattcctgga gctcaggatc
6721 aatctcaaaa ggatataggt aatgcgatgt atttcccaac tgtaagtggc tcattagttt
6781 ctagtgatgc tcagttgttt aataggccct tctggcttca aagagcacag ggtcataata
6841 atggcatcct gtgggctaat cagatgtttg tcacagttgt agacaataca cgaaacacca
6901 atttcagtat atctatttac agtaataaac aaaatgtcac cgacattcca aattatgatt
6961 ctcaaaaatt tagagaatat ttaagacatg tagaggaata tgaaatttct cttatttttac
7021 aattatgtaa agttcctta aaggcagaag tgttagcaca gataaatgca atgaactctt
7081 ctttgctgga ggactggcaa ttaggcttcg tgccaactcc tgacaatcct attcatgaca
7141 catacagata tatcgaatct ctggctacta ggtgccctga taaaaatcct ccaaaagaaa
7201 agccggaccc ttatgatggc ttaagttttt ggactgtaga tatgactgag agactttctt
7261 tagacctgga tcagtattcc ttagggcgca gttcttatt ccaggctggc ctccaacaaa
7321 cgACCGTTAA CGGTacaaca aaatcatcaa gctatagaag ttccataagg gggaccaaaa
        -> E2 bind
7381 gaaaacgcaa aaacTAAatg tACCGAATTT GGTacaatta cctcaacttt gcacagtat
                     <- L1 end
                        -> E2 bind
7441 tcaaggaatg tttgtttact ctgactaagt ctaactctac caaggaaaca gACCGCGCCC
                                                              -> E2 bind
7501 GGTacataaa ggtgagttgg tgccaaatta gtctcacttt gttgccagaa cataccgtgt
7561 tcgtcctaac atgctcggat taggtcgacc gccaaaggac ctttggtttg ccaaatagct
7621 tacagcagct cagttctggc acatttgtgg ACCGATAACG GTaagtctca ttc
                                               -> E2 bind
```

HPV14d

```
LOCUS       HPV14d        7439 bp ds-DNA             VRL        04-OCT-1993
DEFINITION  Human papillomavirus type 14D (HPV-14D), complete genome.
ACCESSION   X74467
SOURCE      Human papillomavirus type 14D DNA.
REFERENCE   1  (bases 1 to 7439)
  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 7439)
  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     Sequence has a deletion of about 274bp (by comparison to closely
            related HPV25) at the HindIII cloning site (pos 622 to 627). The
            plasmid clone sequenced contains at this position a 368bp fragment
            from HPV15.

            HPV-14 was isolated by Kremsdorf et al. in 1984 (J Virol 52:
            1013-8) and subsequently sequenced by Dr. H. Delius.  It has
            been associated with the flat wart-like lesions of EV, a
            multifactorial disease, and infrequently with malignancies.  HPV-14
            is considered to be part of the a$_2$ cluster based on
            phylogenetic analysis.  This cluster includes HPV-19, HPV-25,
            HPV-20 and HPV-21, in addition to HPV-14.  Patients with
            EV tend to have depressed cell mediated immunity.  In roughly
            one-third of EV-associated HPV infection, the sun-exposed flat
            wart-like or macular lesions transform into malignant squamous
            cell carcinomas. Benign wart scrapings tend to be multiply-infected,
            with as many as six different viral types.  However, in contrast,
            EV carcinomas tend to harbor only a few types, specifically HPV-5
            and HPV-8, and less frequently HPV-14, HPV-17, HPV-20 and HPV-47.
            These types are rarely detected in lesions afflicting the general
            population.   A key to host restriction of these viruses may be in
            part due to the unusual organization of the LCR in these viruses.
            The LCR of these viruses is short compared to the viruses in other
            groups and contains two EV-specific regulatory regions: M33 and M29,
            both shown to be involved in protein binding.

            Due to an apparent deletion of approximately 300 bp, the E6 and E7
            ORFs of HPV-14d seem to be disrupted.   Based on its homology with
            other HPV types, the ORF appearing in the sequence from bp 82 to
            744 is similar at its 5' end to the beginning of E6, and at its 3'
            end to the end of E7.  On this basis, we have chosen to assign bp
            196-627 to E6 and bp 628-744 to E7.

BASE COUNT     2337 a   1432 c   1612 g   2058 t
ORIGIN      195 bp upstream from beginning of E6/E7 fused cds
        1 AACGGTaagt tattctgcAC CGGGTGCGGT cactgtatta ctcactatgt ggttgttgtt
E2 bind (end) ->               -> E2 bind
       61 gccaactacc attgctgaTA Gcatgttttt gcctgtaacg ttatcgacac atacatatct
                    E6 orf start ->
      121 atgtatatat atatatat atatatat atatatat atatatacta cagaaaaaac
```

```
 181 agagaatgca gactcATGgc gacaactgac tcttcaacag acagtgcaga tgaaggtcct
                E6 cds ->
 241 tctcctaaga gtaactattg tgatagcaca gaaaccaaat cttctttat agagccacca
 301 ttacctgcaa ctatatttgg cttagcaaac ctattggaaa taccactaga tgattgttta
 361 gtaccttgta acttttgtgg taattttttg actcatttag aagtctgtga atttgatgag
 421 aaaaaactaa gtctaaattg gaaaggtcat tgtgtatttg cttgttgccg tgtatgctgc
 481 acagcaacag caacgtatga gtttaatgaa ttttatgaga gtactgttga aggcagagaa
 541 atagagagtg taacaggcaa atctattttt gatgttgatg tcaggtgcta tacctgcatg
 601 aaattttag attcaattga aaagcttcgc atctttaTAA ctgctacaga atttgctctt
putative deletion of approx. 280 bp /\     E1 orf ->
creating a fused E6/E7 orf                 start
 661 agaaccttcc agaacctgtt atttgaacaa ctgcagctgt tgtgtcctga gtgccgtggg
 721 aactgcaaac ATGgcggatc cTAAaggtag tacatctaaa gacgggttgg atgattggtg
         E1 cds ->            <- E7 end
 781 tattgtggaa gctgaatgta gcgatataga aaatgatttg gaagaattat ttgacagaga
 841 tacagactca gatatttcag aattattaga tgataatgat gacttggacc agggaaattc
 901 tcgggaacta tttcatcaac aagagagtaa ggaaagcgag gagcacttgc aaaaactaaa
 961 acgaaagtac ttgagtcctc aagctatcgc acagcttagt ccgcgacttg aaagtataac
1021 attgtcacct cagcagaagt ctaaacgaag gctctttgca gagcaggaca gcgggttgga
1081 gttaactctt acaaatgaag ctgaagatgt ttcttctgag gtggaggtac cggctctaga
1141 ctctcagccg gttgctgagg cacaaatagg aacagtagac attcattata cagaattatt
1201 acgtgccagc aacaataagg caattcttat ggcaaaattt aaggaggctt ttggggtagg
1261 ctttaatgat ttgacacgtc agtttaaaag ttacaaaacc tgctgtaatc attgggttct
1321 gtctgtatat gcagtgcatg atgatcttct tgaaagctca aagaagttat tgcaacagca
1381 ttgtgattat gtatggatac gtgggatagc tgctatgtca ttattttat tgtgtttcaa
1441 agtgggaaaa aatcgtggga cagtacataa attaatgacc tcaatgttaa atgtgcatga
1501 aaagcaaata ttgtctgagc ctccaaagct acgaaatgtt gctgctgcat tgttctggta
1561 taaaggtgca atggggtcag ggacatttac ttatggtccc taccctgatt ggatggcaca
1621 tcaaactatt gttggccatc aaagtacaga agcaaatgca tttgatatgt ctgttatggt
1681 gcagtgggca tttgataaca attatttaga tgaagctgat atagcctatc aatatgctaa
1741 gttagcacca gaagatagta atgctgtggc ctggcttgcc cataataatc aggccaggtt
1801 tgttagagaa tgtgcatcta tggttagatt ttataaaaaa ggtcaaatga aagaaatgtc
1861 tatgtcagaa tggatacata ctagaataac tgaagtagaa ggagaaggtc attggtcaac
1921 aatagcaaaa tttcttagat atcaacaagt aaactttata atgttttttag ctgctttgaa
1981 agatatgcta cattcagttc ccaaacgtaa ttgtatatta atatatggtc ctccaaatac
2041 tgggaagtca gcatttacca tgtctttaat tcgtgtgtta agaggaaggg tgcttcatt
2101 tgttaattct aaaagccaat tttggctgca accaatgtca gagtgtaaaa tagctttaat
2161 tgatgatgtc acagatccat gttggttgta tatggacact tatttgagga atggccttga
2221 tggtcattat gtttctttag attgcaaaca taaagcaccg atacaaacta aatttcctgc
2281 actattactt acatctaata ttaatgtaca caatgaaata acgtatagat atttgcatag
2341 tagaattaag ggatttgaat ttccaaatcc atttccaatg aaagcagaca atacacctga
2401 atttgaactc actgaccaaa gctggaaatc tttctttaca aggctttgga atcaattaga
2461 gcTGAgtgac caagaagacg agggagacaA TGgagaatct cagcgaccgt ttcaatgctc
E2 orf start ->               E2 cds ->
2521 tgcaagatca gctaatgaac atttaTGAga ctgcagcaaa cacacttgag tcgcaaattg
                           <- E1 end
2581 agcattggca aactcttcga aaagaagctg tgctactata ttttgctagg caaaatggtg
2641 tgacacgact tggataccaa gttgtgccta cattagccat ttcagaagca aaagccaagc
2701 aggccatagg gatggtgctg cagttgcaat cactgcaaaa gtctcagttt ggcagtgaac
2761 catggtcact ggttgatacc agtggagaaa catttagaag tgctccagaa aatcatttca
2821 aaaagggtcc agtatcagta gaggTGAttt ATGataacga taaagacaat gcaaatgctt
                             E4 cds ->
                   E4 orf start ->
2881 atactatgtg gaagcacata tattaccagg atgatgacga acagtggcat aaaagtgcaa
2941 gcggggtcaa ccacacaggc atatattata tgcaaggaac ctttagaaac tactatgttt
3001 tgtttgctga tgatgcaact agatatagta aaactggaca ttgggaagtt aaagttaata
3061 aggaaactgt gtttactcct gtcaccagct ccacccctcc cgagtcacca ggaggacaag
3121 cagactcaaa cacctcctcc aagacccca ccaccgccac tgactccacg tccagactct
3181 cgcccgcaga ttccagaaaa cagtcacaac aagccaacac caaaggaaga aggtacggac
3241 gcagaccgtc cagtaggacc cggcgaacga ccgaaacgcg cagaggcgg agatcgaggt
3301 ccaagtccag gtcgcggtcg aggtcgcggt ctcggctccg atctagatcc cggtcgcaat
3361 cgtctgagcg gcggtctcgg taccgatcaa gatccagatc cagacaaaaa gaagtgtcca
3421 gaatcacaac caccaccaga gggagaggtc gagggtcatc ctccacctcc tccaaacggt
3481 cacaacgggc acgaggaagg ggccgtgggg ggagcagggg gagacggtca tcctccacct
```

```
3541 cccccacctc ctccaaacgg tcacgacgag agtcagagtc ttctaggcag cgtggcatct
3601 ctcctagtga cgtgggaaag tcacttcaat cagttagttc aagaaatca ggaagacttg
3661 gaaggttact ggacgaagct ctcgatcccc cagTAAtctt agtcaggggg gaccctaaca
                                    <- E4 end
3721 cgctacgatg ctttcgcaat agagctaagc aaaagtttac agggctttac agggccttta
3781 gcacggcttg tcgtgggvtg gctggagatg gcactgagcg tctaggcagg tccagaatgc
3841 tcattagctt tttctccttt aaccagagaa gagattttga tcagactgtt aagtacccga
3901 aaggagtgga ccggtcgttt ggctcatttg atagcctaTA Acacccctaa catacTAAca
                                              <- E2 end    L2 orf ->
                                                           start
3961 taatagcttg ctactaacat ctaacatta ttgcattttt gcttttttgtt tgcattattt
4021 taatgctATG gcgcgtgcta ggcgagtcaa gcgtgactct gctactaaca tttacagaac
     L2 cds ->
4081 ctgcaagcaa gcaggcacgt gtcctcctga tgtcattAAT AAAgttgaaa gcacaactat
                                              signal ->
4141 tgctgataaa attttgcagt atggtagtgc tggtgttttt tttggggggtt tgggcataag
4201 cactggaaaa ggtacaggag gtaccacagg ctatgtgcct ttgggagagg gcccagcagt
4261 acgtgttggt ggtgcgccaa caattatcag ACCTGCTCTG GTcccagaca ccattggtcc
                                              -> E2 bind
4321 atcagatatt atacctgtgg acaccttaga tccagtggag cctacgacct cttctattgt
4381 tccactcacg gattccacag gaccagacct tttgcctggc gaggtggaaa ctattgcaga
4441 ggtgcatcct ggcccgtcta ggcctcctac tgacactcct gtcacaacta gtacaggagg
4501 ctccagtgct atattagaag tagcaccgga acctactccg ccctcacgtg ttagggtgac
4561 ccggacacaa tatcataatc cctcctttca agtaattacc gaatccaccc ctaccacagg
4621 tgaaagttca ttagcagaca atatattggt tacctctggt tctggggggac aaactattgg
4681 aggcgctaca cctgaactta tagaacttca agagttacca tctagatatt catttgaaat
4741 cgaagaacca acacccccta gaagaactag taccccatta caaaggatac agacagctat
4801 aagaaggagg ggtgggctta caaataggcg cttagtccaa caagtttctg tagaaaaccc
4861 cttattttta acaagaccat ctagactagt gcaatttcag tttgataatc cagcatttga
4921 ggaggaagta acacaaatat ttgaacaaga tattgaagat tttaatgagc ctccagacag
4981 agattttcta gatgttcaaa ggctgggtag gccccaatat tcagaaactc cagcagggta
5041 tctccgagtt agtcgtcttg ggcaaaggcg gactatacgc actcgttctg gagcacaaat
5101 tgggtctcaa gttcatttt atagagatct aagtagtata aacacagaag atcctattga
5161 gcttcaatta ttaggtcagc attctgggga tgctactatt gtccaaggtc cagttgaaag
5221 cacatttgta gacataaatg tagatgaaaa tccactttct gaggattta gtgcacattc
5281 tgatgacttg cttttagatg aagctaatga agactttagt ggctctcaat tagttgtggg
5341 taatcgacgc tcaacatctt catataccgt ccctcgtttt gaaacaacca gatctgggtc
5401 atattatgca caggatacaa aaggttatta tgtagcttat cctgaggata gggacattag
5461 catggatatt atttatccta ccccagagtt gcctgttgtc attattcaca catatgatac
5521 aagtggtgat ttttacctgc atcctagtct tcacaaaaga ctcaaaagaa aacgaaaata
5581 tttgTAActt tttcttttac agATGgcagt ttggcaagca gctagtggta aggtttacct
           <- L2 end   L1 cds ->
L1 orf start ->
5641 tccaccatct acaccagttg ccagggtcca aagtacggac gaatatgtgc aaaggactaa
5701 catctattat catgcataca gtgacagatt attaactgtt ggtcatccat atttcaatat
5761 atatgacgtg caaagtgcta agataaaagt accaaaagta tctgaaaatc aacataggagt
5821 tttcagacta aagttgccag accctaatcg atttgcatta gctgacatgt ctgtttataa
5881 tccagataaa gaaagactgg tttgggcatg cagaggtata gaaataggca gaggacaacc
5941 tttaggtgta ggtagtgtag acatccatt atttaataag gttggtgata cagaaaatcc
6001 caactcatac aggcaacaag ctaactccac tgatgacaga caaaatgtgt catttgatcc
6061 taagcaactg caaatgttta ataggctg tgcaccttgc atgggggaac attgggatag
6121 ggccttgcca tgtgtagaag ataaaccacc ccctggttct tgccctccaa ttgaattaaa
6181 aaatacagtg attgaagatg gtgacatggc agatataggc tatggaaatt taaattttaa
6241 ggcattacaa gaaaatagat ctgatgtaag tttggatata gttaatgaaa tttgcaaata
6301 tccagacttt ctgaaaatgc aaaatgatgt atatggagat tcctgctttt tttatgcacg
6361 cagggaacaa tgttatgcca gacacttttt tgttagaggg ggtaagacag gagatgacat
6421 accagcagca caagttgatg agggtagcct aaagaatgtt tattacattc caccaatgac
6481 aaatcaacca caaacaata ttggcaatgc catgtatttc ccaactgtca gtggctcatt
6541 ggtatccagt gatgctcaac tgttcaatag ACCATTTTGG TTacagcgcg cacaaggcca
                                              -> E2 bind
6601 caataatggt atttgttggt ttaatcagtt atttgttact gttgtggaca acacacgtaa
6661 cacaaatttt agtatatcag ttagttcaga aacactgag gtatccaaaa ttgacaatta
6721 tacctctcag aaatttcaag aatattaag acatgtagaa gaatatgaaa tgtctctaat
6781 tttacaacta tgtaaaatac ctttaacagc tgaagtctta gctcaaatta atgcaatgaa
```

```
6841 ttctaatatt ttagaggagt ggcaattagg atttgtacct gcaccagaca atcctattca
6901 tgatacatac agatatattg agtctgcagc gactaggtgt cctgataaaa atcctcctaa
6961 agaaagagaa gatccttata aaaactttaa cttttggaat gtagatttaa cagagagact
7021 atctttagac ctagatcaat attctcttgg gagaaaattt ttatttcagg caggtttgca
7081 gcaatcgACC GTTAACGGTa caaaaacagt ttcgactagg ggatccatca agggtattaa
                -> E2 bind
7141 acgaaaacgc aagaatTAGa cattatcgat ttcggtgcaa taaagtcaac ttttacacag
                       <- L1 end
7201 tattcaagga atgtttattc actctgacta agcaaatatg agccgcgccc gatacataaa
7261 ggtgccaaat gaggtgagtt gtttgccaga agaggtcaga gccaactcag gtttgcgcca
7321 gatcagatac agcgcgagcc gcgttggatc aagctacatc gtctgaacac gcaaaagact
7381 caaggaaatg taagtgtgcc agtctattgt gttcgaattt ggcaaagttg aagACCGTT
                                                            -> E2 bind
                                                            (start)
```

```
LOCUS       HPV15         7412 bp ds-DNA             VRL        04-OCT-1993
DEFINITION  Human papillomavirus type 15 (HPV-15), complete genome.
ACCESSION   X74468
SOURCE      Human papillomavirus type 15 DNA.
REFERENCE   1  (bases 1 to 7412)
  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 7412)
  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-15 was isolated by Kremsdorf et al. in 1984 (J Virol 52:
            1013-8) and subsequently sequenced by Dr. H. Delius.  It has
            been associated with the benign flat wart-like lesions of EV,
            a multifactorial disease.  HPV-15 is considered to be part of
            the b cluster based on phylogenetic analysis.  This
            cluster includes HPV-15, HPV-17 and HPV-9.  Patients with EV
            tend to have depressed cell mediated immunity.  In roughly
            one-third of EV-associated HPV infection, the sun-exposed flat
            wart-like or macular lesions transform into malignant squamous
            cell carcinomas. Benign wart scrapings tend to be multiply-infected,
            with as many as six different viral types.  However, in contrast,
            EV carcinomas tend to harbor only a few types, specifically HPV-5
            and HPV-8, and less frequently HPV-14, HPV-17, HPV-20 and HPV-47.
            These types are rarely detected in lesions afflicting the general
            population.   A key to host restriction of these viruses may be in
            part due to the unusual organization of the LCR in these viruses.
            The LCR of these viruses is short compared to the viruses in other
            groups and contains two EV-specific regulatory regions: M33 and
            M29, both shown to be involved in protein binding.


BASE COUNT     2374 a   1321 c   1611 g   2106 t
ORIGIN      199 bp upstream from beginning of E6 cds
        1 caagtaactt ggcagaacat tttcttggaa gacagcACCG ATAACGGTaa gattatatct
                                                       -> E2 bind
       61 ttgaACCGTA GGCGGTtctt tctgattggt ttggctgata gtagttaaca acaatcactc
             -> E2 bind
      121 tttataaata tatgtaACCG CCTGCGTTac tttatttaat ctacatacaa tatgcTGAgt
                        -> E2 bind                        E6 orf start ->
      181 aaactattta gagagctatA TGgataggcc aaagcctttt tctgtgcagc agcttgcaga
                  E6 cds ->
      241 cactctgtgt ataccttttag tagatatatt attgccttgt agattttgtc agagattttt
      301 aacatatata gaattagtaa gtttgaatcg taaaggtctg cagttaattt ggactgagga
      361 agattttgtt tttgcctgtt gttctagttg tgcatttgct acagcgcagt ttgaattttc
      421 taactttttat gaacagtcgg tgtgtagttg ggaaatagag atagtagaac agaagcctgt
      481 tggagatatt attattcgct gcaaattttg tctgaagaaa ttagatttaa ttgaaaagtt
      541 agatatttgt tacaaagagg agcaattcca caaggttaga cgcaattgga aaggattgtg
      601 tagacattgT AGggcgatag aATGAttggg aaagaagcta ctataccaga tatagtgctt
          E7 orf start -> E7 cds ->
                                       <- E6 end
      661 gagctgcaag agcttgtcca gcccactgac ctgcattgct acgaagagtt aagtgaagaa
      721 gagacagagg aggagccacg atttattcct tacaagattg tagttccgtg ttgctttttgt
      781 gattccaagc ttcgacttaT AGtggttgca actccatttg gaattcgctc acaacaagac
                  E1 orf start ->
      841 ttattattgg aagaagttaa gttggtgtgt ccagggtgtc gagagaagct tcgccATGtc
                                                                E1 cds ->
      901 TGAtgacaaa ggtacatatg atcctaaaga aggctgtagt gattggtttg ttctagaagc
          <- E7 end
      961 agaatgctct gatgctagtt tagatggtga tttggaaaag ttatttgaag aaggtacaga
     1021 tactgacatt tctgacttaa tagataatga ggacactgta caggggaact cccgcgaatt
     1081 attatgccag caagaaagtg aggaaagcga gcaacaaata cactggctaa aacgaaagta
```

```
1141 tatcagttca caagaggttt tgcagcttag ccctcgcctg cagtgtatat ctatttcgcc
1201 acagcataag tctaaaagga gattatttga caagacagc ggactagaac tatcatttaa
1261 tgaagctcaa gatttactc agcagacttt ggaggtaccg gcgaccgatg ttgtgccgca
1321 gggtgccaag ggactgggca ttgttaaaga tcttcttaaa tgtaacaatg tgaaagcaat
1381 gttattagct aaatttaaag aagcgtttgg agtggggttt atggaattaa ctagacaata
1441 taaaagcagc aaaacatgct gtagagactg ggtactgact gtttatgctg tacaagatga
1501 gctgttagaa agttctaaac aattgttgat tcaacactgt gcatatattt ggttacatca
1561 aatacccct atgtgcttat atttattgtg ctttaatgtt ggtaaaagta gagaaacagt
1621 attaagatta cttacgaatt tgttacaagt atctgaaata caaataatag cagaaccacc
1681 aaagcttcga agtacactgt cagcattatt ttggtataaa ggaagtatga atccaaatgt
1741 ttatgctcat ggagaatatc ctgagtggat aatgacacaa acaatgataa atcaccaaac
1801 agcagaagct acacagtttg atttatctac tatggtacaa tatgcatatg ataatgaatt
1861 gtcagaagaa gcagaaattg cttggcatta tgcaaaatta gcagatacag atgcaaacgc
1921 gagagcattt ttacagcata acagtcaggc aagacttgtc aaagactgtg caataatggt
1981 tagacactat agacggggag aaatgaaaga aatgtctatg tcatcatgga tacataaaaa
2041 gttattggtt gttgaaggag aaggacattg tcagatatt gtaaagtttg tcagatacca
2101 agatataaat tttatacaat tttttagattc atttaaaagc tttttacata atactcctaa
2161 aaaaagttgt atgttaatat atggtccacc tgacacagga aaatccatgt ttactatgtc
2221 attaataaaa gtcttaaagg gtaaagtttt gtcatttgca aattataaaa gtacattttg
2281 gcttcaacct gtggcagata caaaaaatagc tttaatagat gatgtaactt atgtttgttg
2341 ggattatata gatcaatatt taagaaatgc attggacggt ggtgttgttt gtttagatat
2401 gaaacacagg gcgccatgtc aaataaggtt tccaccatta atgctaactt ctaacattga
2461 tatcatgaaa gaagaaaggt ataaatattt acgcagtaga gtgcaagctt ttgcatttcc
2521 acataagttt cctttgata gtgataataa tccacaattt aaacttactg accaaagctg
2581 gaaatctttt tttgaaaggc tttggagaca gtTAGagctc agtgatcaag aagacgaggg
                    E2 orf start ->
2641 agacgATGga tactctcagc gaacgtttca atgcactaca agagaatcta atggacattt
  E2 cds ->
2701 aTGAgtcagg tcgagacgac atagaaactc aaatattgca ctggcaatat ttgaggcaag
         <- E1 end
2761 aacaagtatt attctatttt gccaggaaac atggagttat gcgtttagat atcagcctgt
                                            /\ putative
                                         deletion causing premature
                                         termination of E2 cds
2821 acccccttTA Gccaccagtg agaccaaagc gaaagatgct attggtatgg tgattctgtt
              <- premature termination of E2 cds
2881 acaaagtttg cagaagtcag catatggcaa ggagccatgg acactaacac agactagttt
2941 ggagactgtg agaagtgcac ctgcaaactg ctttaaaaaa gggccacaga atattgaagt
3001 tatgtttgat aaggatcctg aaaacattat ggtatacact gtttggacat acatttatta
3061 ccagacttTA GATGacacat ggaacaaggt ggaaggaaaa attgactatc atggcgcata
  E4 orf and cds ->
3121 ttatttggaa ggaactctta aagtttatta catacagttt gaagttgatg cagccaggtt
3181 tggcaaaact ggaatctggg aggtgcatgt taatgaggac actatctttg ctcctgttac
3241 tagctcttcg ccggcagctg gagaaggggc aacctccatc gactccgcac ccgaatcgcc
3301 ggccaacaga cagctttctt ccacctccgt gtcctccaga aaacggacac caccacgaac
3361 cgaagccaga cgctacaacc gaaaagaatc tagccctaca accaccacca cccggaggca
3421 gaaaagacaa ggacaaagac aagaagacac agcaaggcga tcaaggtcca cctcaagggg
3481 gagacaagaa atctccaggg agggaacca gcgccgacgg cgacgatccc gagaaacctc
3541 catctcccccc gcctggggaa gggagggag aagtagaagg gggcccacaa caaggtccca
3601 atcgaagtcc ctctcacgat cccgatcccg atccaagtcg agatcacgag ggtcttctcc
3661 acggggtggc atctcgcctg cagacgtggg aagctcagtt cgatcacttg gtagaaaaca
3721 tactgggcga cttgaaagat tactggaaga ggctagggac cccccagTAA tcttgctgcg
                                                   <- E4 end
3781 cggtgatgct aacaaattaa aatgctttcg ctttagggca aagaaaaagt atcaggattt
3841 agtaaaatac tatagcacca cgtggtcctg ggtaggggt acaagtaatg atagaattgg
3901 acgctcacga ttgttactgg cattctcttc caacaccgaa agagagctct ttatcaaaat
3961 aatgaaactg ccaccaggcg ttgattggtc gctaggatat ttagatgatt taTGAttttt
                                                       <- probable
                                                       E2 end
4021 gtgctttttta atcaactaac agtagtgttt ttttattgct tttgctacTA Acactatact
                                            L2 orf start ->
4081 aacattcccA TGgcccgtgc acgcagagta aaacgtgcat ctgtaactga catttacagg
      L2 cds ->
4141 gggtgcaaac aagcaggtac ctgcccccct gatgttcttA ATAAAgtgga acaaacaact
```

```
                                       signal ->
      4201 attgcagata aaattttgaa atatggaagt gctgctgtat tttttggtgg gctgggaatt
      4261 ggtacaggcc gtggttcagg tggtgctaca ggttatgttc ctttgggaga aggccctggt
      4321 gtgcgcgtag gaggcacccc taccattgtt cgccctgggg tcacacctga acttattggt
      4381 ccagcggatg taataccaat tgatacagtc acaccaattg accccgcagc acctagtatt
      4441 gtcacaataa ctgacagcag cgctgttgac cttttacctg aattggagac aattgcagaa
      4501 atacatcctg taccaacaga taatgtagat attgatactc ctgttgttac aggaggtcgg
      4561 gattccagcg ctattttaga agttgctgat cccagtcctc ctgtacgaac aagagtgtct
      4621 cgcacacaat atcataaccc atcatttcaa attattactg agtctacacc tttgtctggc
      4681 gaatctgcac tagcagatca cgttattgtt tttgaaggta gtggaggtca aaatataggc
      4741 ggttctcgca gtgctgcttt ggatgcagca caggaaagtt ttgaaatgca aacatggcct
      4801 agtagataca gttttgaaat acaagaaggt acaccaccaa gatctagcac tcctgttcaa
      4861 agagcagtac aatccctttc tagtcttaga agggctttat ataacagacg gctgactgaa
      4921 caagtagctg taacagaccc tttattcttg ggacgaccct cacgcttggt gcaatttcaa
      4981 tttgacaatc caacatttga agaagaagta acacaaactt ttgaaagaga tgttgaagca
      5041 tttgaagagc ctccagatag gcaatttcta gatgtagttc gcctaggaag gcctacttat
      5101 tctgaaacac cacaaggtta tgtgcgtgta agtagacttg gtagacgggc cacaattaga
      5161 ACCAGAAGTG GTgcacaagt tggtgctcaa gtacatttct atagggattt aagtacaatt
           -> E2 bind
      5221 gattctgaag cattagaaat gcaattacta ggagaacatt caggtgatag taccattgtt
      5281 caggctccta tggaaagttc atttatagat ataaatattg atgagcctga ttcattacat
      5341 gtgggcctac aggacagtac tgaagcagat gacattgatt acaattctgc tgatcttctt
      5401 ttagaagata atatagaaga ttttagtggt tctcatttgg tgtttggcaa tacacggcgc
      5461 agcactacaa catatacagt acctagattt gaatcaccta gaaatactgg gttttacata
      5521 caagatgtgc atgggtataa tgtagcctat ccggagtccc gtgatactac cgaaataata
      5581 ttaccacaat ctgacacgcc aactgtagtt ataaattttg aagaggcagg tggagactat
      5641 tatttacatc caagcttaaa aacacgTAAa cgaaaacgca aatatttgTA Attgttttac
                          L1 orf start ->                      <- L2 end
      5701 agATGacatt gtggctaccg acgacgggta aagtatattt gccaccaaca ccacctgtag
      L1 cds ->
      5761 cacgtgtaca aagcaccgat gaatatgtgg agagaactaa tgtatttttac catgcaatga
      5821 gtgaccgtct gttaacagta gggcatccct actttgatgt tagatctgtt aatggaggta
      5881 gcatagaagt tcctaaagtg tctggaaatc aatatagagc atttagggtt acttttccag
      5941 atcctaatag atttgcatta gcagacatgt ctgtctataa tccagaaaaa gaaaggttgg
      6001 tttgggcctg tgtaggcctt gaaataggta gaggacaacc attaggagtt ggtacttcag
      6061 gccatccttt attcaacaaa gtaaaagata cagagaataa cagtaattat caaggcaact
      6121 ctactgatga caggcaaaat acatcttttg acccaaagca ggttcaaatg tttgtagtag
      6181 gctgtgttcc atgtttaggc gaacactggg atagagctct tgtttgtgaa tcagagagaa
      6241 ataatcaggc gggaaaatgt cctccttttgg aacttaaaaa tacagttatt gaagatggcg
      6301 acatgtttga tataggtttt ggtaacatta ataacaaggc cttatcagtt actaagtcag
      6361 atgtgagtct ggatatagta aatgaaactt gcaaatatcc agattttta actatggcaa
      6421 atgatgtgta tggagacgct tgttttttt ttgcaagacg agaacagtgc tatgctagac
      6481 attactttgt cagaggaggt gcagtaggtg atgctcttcc tgatgcagct gtcaatcaag
      6541 atcataattt ttatttacca gcacaatcaa cccaacaaca aaataactta gcaaattcta
      6601 cttactttcc cacagttagt ggttcttag tgacatctga tgctcagctg tttaatagAC
                                                                       ->
      6661 CGTTTTGGTT aagaagagct caagggcata caatggcat actttggggt aatcagatgt
           E2 bind
      6721 ttattactgt tgcagataac acaaggaata caaattttac tattagtgtt ACCTCTGATG
                                                                    -> E2 bind
      6781 GTaatgccat aaatgaatat aattcacaaa atatcagaga attttttaaga catgtggaag
      6841 aatatcagtt atctattatt ttgcaattgt gtaaaatacc tttaaaagct gaggtattaa
      6901 cacaaattaa tgctatgaat tcaggtattt tagaagactg gcaactaggg tttgttccta
      6961 caccagacaa cgctgtacaa gatatttata gatatattga ctctaaggca actaaatgtc
      7021 ctgatgctgt acaaccaaag gacaaagaag acccattgg aaagtataca ttttggaatg
      7081 tagatttaac agaaaagtta tctttagatt tggatcaata tcctttggga agaaaattta
      7141 tatttcaagc aggtttgcaa cgtcgcccca gaactattaa atcctctgta aaagtttcta
      7201 aaggtactaa acgcaaacgt acaTGACCGA TTTCGGTcgc taataaacaa gtaaaccaat
                               <- L1 end
                                -> E2 bind
      7261 aaggtatgtg aagtattttt taccatgttc gtgactaaac cgcataggtc attgccaaca
      7321 ACCGCACCCG GTtaatcaga tataaaatgc acctggtgcg attttatcac cgctttttgtg
           -> E2 bind
      7381 gaagcacccg aggcgcccgc cagaactgct gc
```

```
LOCUS       HPV17         7427 bp ds-DNA         VRL       04-OCT-1993
DEFINITION  Human papillomavirus type 17 (HPV-17), complete genome.
ACCESSION   X74469
SOURCE      Human papillomavirus type 17 DNA.
REFERENCE   1  (bases 1 to 7427)
  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 7427)
  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-17 was isolated by Kremsdorf et al. in 1984 (J Virol 52:
            1013-8) and subsequently sequenced by Dr. H. Delius.  It has
            been associated with the benign macular lesions of EV, a
            multifactorial disease, and subsequently from squamous cell
            carcinomas and the malignant melanoma of an immunosuppressed
            patient.  HPV-17 is considered to be part of the b cluster
            based on phylogenetic analysis.  This cluster includes HPV-15,
            HPV-17 and HPV-9.  Patients with EV tend to have depressed cell
            mediated immunity.  In roughly one-third of EV-associated HPV
            infection, the sun-exposed flat wart-like or macular lesions
            transform into malignant squamous cell carcinomas. Benign wart
            scrapings tend to be multiply-infected, with as many as six
            different viral types.  However, in contrast, EV carcinomas tend
            to harbor only a few types, specifically HPV-5 and HPV-8, and less
            frequently HPV-14, HPV-17, HPV-20 and HPV-47.  These types are
            rarely detected in lesions afflicting the general population.
            A key to host restriction of these viruses may be in part due to
            the unusual organization of the LCR in these viruses.  The LCR of
            these viruses is short compared to the viruses in other groups and
            contains two EV-specific regulatory regions: M33 and M29, both
            shown to be involved in protein binding.

BASE COUNT     2329 a   1366 c   1670 g   2062 t
ORIGIN      199 bp upstream from beginning of E6 cds
        1 ctggagccaa ggtttttggca gaacaccatc ttggaacagc ACCGATAACG GTaagattat
                                                        -> E2 bind
       61 atcttggaAC CGTAGGCGGT actttctgat tggtttggct gatagtagtt aacaacaatc
                     -> E2 bind
      121 ttctctcata aatatatgtg ACCGCCTTCG TTaccttaaa tgatctacat acaatatgag
                                -> E2 bind
      181 agctcttact TAAgagcatA TGgataggcc aaaacctcaa acagtgaggg agcttgctga
                        E6 cds ->
          E6 orf start ->
      241 taccttgtgt attccattag tggatatttt attaccttgc agattttgta ataggttttt
      301 agcttacata gaattggtgg cgtttgattt aaaaggtttg cagttaattt ggactgaaga
      361 agattttgtg tttgcctgtt gcagtagttg tgcgtatgct acagcacagt atgaattttc
      421 taagtttttat gaacaatcag tgagtggaag ggagttagag gaaatagagc acaagccaat
      481 aggggaaata cctattcgct gcaagttttg tttaaagaaa ttggatttac tagagaagtt
      541 agacacttgt tatagacatc agcagtttca taaggtTAGa cgcaattgga aaggcttgtg
                                       E7 orf start ->
      601 cagacattgt gggtcgatag gATGAttggg aaagaagcta caataccaga tatagtgctt
                        E7 cds ->
                                  <- E6 end
      661 gagctgcaac agcttgtcca gcccactgac ctgcattgct acgaagagtt aagtgaagaa
      721 gagacagaga cagaggagga gcctcgtcgt ataccataca agattgtagc tccgtgctgc
      781 ttttgtgtt ctaagctacg gcttattgtt cttgcaacgc acgctggaat tcgttcacaa
      841 gaggagcttt tatTAGgtga agtacagttg gtgtgtccta actgcagaga gaagcttcgc
             E1 orf start ->
      901 cATGacTGAc gacaacaaag gtaccaaatt tgatcctaaa gaaggatgta gtcagtggtg
     E1 cds ->
                    <- E7 end
```

```
 961 tatactagaa gcagaatgtt ctgacaatag tttagatggt gatttggaaa agttatttga
1021 agaaggtaca gatactgaaa tttctgactt aatagatgat gaggacatta tacagggaaa
1081 ctcccgcgaa ttgttatgcc agcaagaaag tgaggaaagc gagcaacaga tacaattgct
1141 aaaacgaaag tatttgagtt cacaagaggt tttgcagctc agtccgcgcc tgcagtctat
1201 cactatatcg ccacagcata agtctaaaag gagattattt gaacgagaca gcggactaga
1261 actgtcattt aatgaagctg aagatcttac tcagcagact ttggaggtgc aggaggtatc
1321 ggcgaccggt tctgtaccgg cagaacaggg tgtcaaggga ttgggaattg ttaaagacct
1381 tttaaaatgt agtaatgtga aagctatgtt attggccaaa ttcaaagaag catttggagt
1441 gggatatatg gatttaacca gacagtataa aagtagtaag acatgctgta gagattgggt
1501 agttacattg tatgcagtac aggatggagtt gatagaaagc tccaaacaat tgctgctgca
1561 acattgtgct tatatatggc tacaacatat gtcacctatg tgtttatatt tattatgttt
1621 taatgtcgga aaaagtaggg aaactgtatc acgattgctt atgaatattc tgcaagtagc
1681 agaggtacaa atgttagcag aacctccaaa attgagaagc atgttgtcag cattattttg
1741 gtataaagga agtatgaatc caaatgtgta tgcccacggt gaatatcctg agtggatttt
1801 aacacaaact atgattaatc atcaaacagc acaggcaaca caattcgatc tatctaccat
1861 gatacaattt gcttatgata cgaatacct tcaagaagat gaaatagctt atcattatgc
1921 taaattagca gatacagatg ctaatgcacg agcatttta cagcataata gtcaagcacg
1981 gtttgtaaaa gaatgtgcaa taatggtgag acattataag cgtggagaaa tgaaggaaat
2041 gagcatttct acgtgggtac atagaaaatt attagttgtt gaaggagatg gacattggtc
2101 tgatatagta aaatttatta gatatcagga cattaatttt attaggtttt tagatatatt
2161 taaatcattt ctgcacaata aaccaaaaaa aaactgtata ttaattcatg gcccaccaga
2221 tacaggtaaa tctatgttta caatgtctct tataaaagtg ttgaaaggca aagtgttatc
2281 atttgcaaat tgtagaagca atttctggct tcagccatta gcagacacta agcttgcatt
2341 aattgatgat gtgacatttg tatgttggga ctatatagat caatatttaa gaaatggatt
2401 ggatggtaat gttgtgtgtt tagatttgaa acatagagcg ccatgtcaaa ttaaatttcc
2461 accattacta ttaacatcta atattgatgt aatgaaagaa gacaaatata gatatttaca
2521 cagtagaatt caaagctttg cttttccaaa taagtttccg tttgataata caatatgcc
2581 acaatttcga cttactgacc aaagctggaa atcttttttt gaaaggcttt ggcatcagtt
2641 agatcTGAgt gatcaagaag aagagggaga cgATGgacaa tctcagcgaa cgtttcaatg
     E2 orf start ->                    E2 cds ->
2701 tactgcaaga gaacctaatg gacatttaTG Agtcaggtca agaagatata gagactcaaa
                                   <- E1 end
2761 taaaacactg gcaattatta agacaggaac aagtactgtt ttactatgcc agaaaaaatg
2821 gagtgatgcg tgtaggttat caacctgtgc ctcctttagc caccagtgag gctaaagcaa
2881 aggatgctat tggcatggtt ttgttattgc agagcttgca aaaatcaccg tatggcaaag
2941 agccatggac gctaacacaa actagtttgg agactgtgca agtccaccc gcaaactgtt
3001 ttaaaaaggg tcctcaaaac attgaagtta tgtttgacaa tgaccctgaa aatcttatgt
3061 catatacagt gtggtcattt atttattacc aaaaatttaga tgacacctgg aataaagttg
3121 agggccgtgt tgactatcat ggtgcatatt atatggaagg ctctcTAAaa gtgtattata
                                   E4 orf start ->
3181 ttcaatttga agtggATGct gccaggtttg gcaaaactgg acgttgggaa gtgcatgtta
              E4 cds ->
3241 atgaggacac tatctttgct cctgttacta gctcttcgcc ggcagctgga gaagggaccg
3301 acgcgtccc catcaacgcc gcatcccggt cgtcaccagc aagggggactc tctgccacct
3361 ccgtgtccac ccgaaccaca caacggacat caccacggcg atacaggcgg aaaagcgtcta
3421 gccctacagc caccaccacc cggcacaaaa gacaagacat cagacgatca aggtccacct
3481 cacggggggag acaagcaatc tccaggggag gggagcgacg ccagcggaga cgagaacgct
3541 cctactcccg agactcctca agatccccca acaggggaag gggagggagc agtggggggc
3601 ccacaacacg gtcccagtcg cggtccctct cccgatcccg atcgcggtca cgatcgcgat
3661 ccagagggtc ttctgccggg ggtggcgttg cgcctgagca agtgggaaaa tcagttcgat
3721 cagttggtag aaaccctggt gggcgactta caagattatt ggaagaagct agggatcccc
3781 cagTAActttt gctgcgcggc gaagctaaca aactgaaatg ctttcgatat agagcaaaaa
          <- E4 end
3841 agcgatatgg cagtttagtt aaatattaca gcactacatg gtcatggggtg ggtgcaaata
3901 ctaatgacag aataggtaga tcaagaatgt tactagcatt taacacatat gatgaaagag
3961 aattgtttat ccaaaaaatg aagctaccac caggtgttga ttggtcacta ggacatctag
4021 atgatttaTA Ggcattactt ttttaacatt actaaccttg cTAAagcttg cttttgctac
              <- E2 end                         L2 orf start ->
4081 taaacacttt taacgttccc ATGgctcgct caagacgcat aaagcgtgcc tctgtaactg
                  L2 cds ->
4141 acatctacag ggggttgcaag caggctggta cttgccccccc tgatgttatt AATAAAgtgg
                                                            signal ->
4201 aacagactac aatagcagat aaaatttttaa aatatatggtag ttctggtgtt tttttttggtg
4261 ggctgggcat tagcacaggt cgtggcacgg gtggggcaac aggttatttc cctttgggtg
```

```
4321 aagggcctgg ggtacgcgta ggtggcgccc ccactatagt tcgccctggg gtcatacctg
4381 aactcattgg cccagcggat gtaataccta ttgatacagt cactccaatt gaccccgcag
4441 cacctagtat tgttacaatc acagacagca gtgctgttga cctattaccc actgaattgg
4501 aaaccattgc agaaatacat cctgtgccta cagataattt agatattgac actcctgttg
4561 tttcaggagg cagggattcc agtgctgttt tagaggttgc tgatcctagt ccccctgtaa
4621 gaacaagggt gtctcgaaca cagtatcaca acccatcatt tcaagtaatt actgaatcta
4681 cacctttatc tggagaatca gctatgcgg atcatgtttt agtgttcgaa ggttttggtg
4741 gacaaaacat aggaggttcc aggaatgcag ccattgatac agcacaggag agctttgaaa
4801 tgcaatcctg gcctagcaga tatagctttg aattagaaga aggcacacct cctagaacaa
4861 gtactccagt tcaacgtgca gtagaatcac tatcaagctt aagaagagct ttatacaata
4921 gacgattgac tgaacaagtt gcagtgacag atccactttt tttaagtagg ccttcacgct
4981 tggtccaatt tcagttcgat aatcctgcct ttgaggaaga agttacacag ctgtttgaaa
5041 gagatattga agcagtggag gaacctcctg atagacagtt tttagatgtt gtgcgcctag
5101 gaaggcctac atattctgaa acacctcagg gttatttacg agtcagtaga ctaggtagaa
5161 gagccagcat tcgtactcgc agtggagcac aagtaggagc tcaggtacat ttttatagag
5221 atgttagcac catcgattca gatgccttag aaatgcagtt attgggggaa cattctggtg
5281 acactaccat agttcagggt cctgtagaaa gttcctttgt agacattaat attgatgaac
5341 cagggccttt gaatgtaggc atccaagaat caccactggc tgacactata gaagaagatt
5401 tcaattctgc agatttgtta ctggaagatg ctgtagatga ctttagtggg tctcagctgg
5461 tatttggcaa tcctcgccgc agcacaacat ctgtaactgt cccccggttt gaaacaccta
5521 gggacactgg ctttttacata catgacactc agggatacac agtagcatat ccagagtcac
5581 gtgacaccac tgaaataatt cttccacatc ctgatacacc aactgtagta attaaatttg
5641 cagaagcagg aggcagattt ttatttacac ccTAGtttaa agaaacgaaa aagaaaacga
                        L1 orf start ->
                                        <- L2 end
5701 aaatatttgt aattgtttttg cagATGacat tgtggctgcc aacgaccggt aaagtatact
                        L1 cds ->
5761 tgcctccaac accaccagta gcccgagtac aaagcacgga tgagtatgtg gaaagaacaa
5821 atatttttta ccatgctatg agtgatcgtc tcctaactgt gggacaccca tttttatgatg
5881 taagatctac tgatggatta agaatagaag tccctaaagt ttctggaaat cagtatagag
5941 ccttcagagt cacattacca gatcctaata agtttgcctt ggcagatatg tcagtttaca
6001 atcctgaaaa ggaaagatta gtttgggcat gtgcaggcct tgagatagga cgaggacagc
6061 cattaggtgt aggcactaca ggacatccct tgtttaataa gttaagagac actgaaaata
6121 acagtagcta tcaaggtgga tctactgatg atagacaaaa cacgtcattt gaccctaaac
6181 aagtgcagat gtttgttgta ggctgtgtac cttgtattgg agaacattgg gacagggctc
6241 ctgtatgtga aaatgaacaa aacaatcaaa caggcctgtg tccaccattg gaattaaaaa
6301 acactgttat cgaagatggt gacatggttg acataggctt tggaaacatt aataacaaag
6361 tgctttcatt taataaatca gatgtaagtt tagatatagt taatgaaaca tgcaaatatc
6421 ctgattttttt aagcatggca aatgatgttt atggtgatgc atgtttcttt ttcgccagac
6481 gagagcagtg ctatgctagg cattattttg tgaggggagg taatgtaggt gatgctgttc
6541 cagatggttc tgtaaatcag gatcacaaat tttatttacc agctcaaact ggccaacaac
6601 agcgcacttt gggtaattcc acttattttc caactgtaag tggttcttta gtaacatctg
6661 atgcccagct ttttaatagA CCATTCTGGT Tacgtagagc acaaggacac aataatggta
                        -> E2 bind
6721 ttttatgggg gaatcagata tttgtgactg tagctgacaa cactaggaac acaaattttt
6781 ctattagcgt gtctacagaa gctgggggctg ttacagaata taattctcaa aatatcagag
6841 agtatttaag acatgtagaa gagtatgatc tatcttttat tttacaatta tgtaaaatac
6901 ctttaaaagc tgaagtttta actcaaatta atgcaatgaa ctcaggaatt ctggaagact
6961 ggcagttagg atttgtgcct acaccagata acccagtgca tgtatatatat aggtacatta
7021 attctaaagc cacaaaatgt cctgatgcag ttgtagaga agaaaaagaa gacccttttg
7081 caaagtatac attttggaat gtaaatctta ctgagaaatt atcattagat ttagatcagt
7141 atcccctagg acgaaaattt atttttcagt caggtttgca ggcaaggccc agaactattc
7201 ggacctctgt aaaagtgccc aagggtatta aacgcaaacg ttcaTGACCG CTTTCGGTct
                                                <- L1 end
                                                -> E2 bind
7261 ctcaataaac aaataaacca aaatggtatg tgaagtattt tttaccatgt tcgtgactaa
7321 accgtcttcg tcaacgccag aaACCGCACC CGGTaatca gattataaat gcacctggtg
                        -> E2 bind
7381 cgattttatc actgcttttg tggaagcacc gcaggcgccc gccgaaa
```

```
LOCUS       HPV19          7685 bp ds-DNA            VRL       04-OCT-1993
DEFINITION  Human papillomavirus type 19 (HPV-19), complete genome.
ACCESSION   X74470
SOURCE      Human papillomavirus type 19 DNA.
REFERENCE   1  (bases 1 to 7685)
  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 7685)
  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-19 was isolated by Gassenmaier et al. in 1984 (J Virol 52:
            1019-23) and subsequently sequenced by Dr. H. Delius.  It has
            been associated with the benign macular lesions of EV, a
            multifactorial disease.  HPV-19 is considered to be part of
            the a$_2$ cluster based on phylogenetic analysis.  This cluster
            includes HPV-14, HPV-25, HPV-20 and HPV-21, in addition
            to HPV-19.  Patients with EV tend to have depressed cell mediated
            immunity.  In roughly one-third of EV-associated HPV infection,
            the sun-exposed flat wart-like or macular lesions transform into
            malignant squamous cell carcinomas. Benign wart scrapings tend to
            be multiply-infected, with as many as six different viral types.
            However, in contrast, EV carcinomas tend to harbor only a few
            types, specifically HPV-5 and HPV-8, and less frequently HPV-14,
            HPV-17, HPV-20 and HPV-47. These types are rarely detected in
            lesions afflicting the general population.   A key to host
            restriction of these viruses may be in part due to the unusual
            organization of the LCR in these viruses.  The LCR of these viruses
            is short compared to the viruses in other groups and contains two
            EV-specific regulatory regions: M33 and M29, both shown to be
            involved in protein binding.

BASE COUNT     2424 a   1460 c   1663 g   2138 t
ORIGIN     199 bp upstream from beginning of E6 cds
        1 GGTaagttta ttgatacggg cgcggttaga agttactcat tcggtgcttg ttgttgccaa
   E2 bind <-
       61 caatcagcgt tatgaacttg tttctgcctg tatcggtatc gacacaggta ttatatatat
      121 atatatatat atatatatat atatatatat atatatatat atatatacac acacagatac
      181 attttgcagc tgcaaacttA TGgctaacgc acaggctaca gaagaagaga tagaaattgt
                      E6 cds ->
      241 agaagaggga actactgcac cacaggtcac agagccacca ttaccagcaa caattgctgg
      301 attagcagca ttgctagaaa taccgttgga tgactgttta gtgccttgta atttctgtgg
      361 caagttttta tcacatttag aagcgtgcga atttgatgat aaaagactta gtttgatttg
      421 gaaaggtcat cttgtgtatg cttgctgtcg ctggtgttgc acagcaactg caacatttga
      481 atttaatgag ttctatgagc atactgtaac aggtagagaa attgagtttg taacaggtaa
      541 atctgtcttt gacattgatg ttagatgtca aaattgcatg agatatcttg attcaattga
      601 aaagcttgat atctgtggaa gaagacttcc ttttcataaa gTAAgagact cttggaaagg
                                                  E7 orf start ->
      661 gatctgtagg ctgtgtaagc atttctataA TGattggTAA agaggtgata ttgcaagaca
                          E7 cds ->          <- E6 end
      721 ttgtattaga attaagtgag ttgcagcctg aggtacaacc agttgacctg ttttgtgaag
      781 aggagttacc gaccgaacag caggaaacag aggaggagcc tgctattgaa agatctgcgt
      841 acaaagttgt tgtactttgt ggctgctgca aggtgaagct tcgcatcttt gtgaaagcca
      901 cgcaatttgg tattagaacc ctacaggaca tccTGAttga agaattgcaa ctgttgtgcc
                          E1 orf start ->
      961 cggagtgccg tgggaactgc aatcATGgcg gagtcTAAag gtagtacatc taaagaaggg
                      E1 cds ->          <- E7 end
     1021 tttggtgatt ggtgtatttt ggaagctgaa tgtagtgatg tagaaaatga tttggaaaaa
     1081 ttatttgatg aagatacaga ctcagatatt tcagacttat tagatgataa tgacttggag
     1141 caggtaact ctcgggaact atttcatcaa caagdgtc aggaaagcga agagcatttg
     1201 caaaaactaa aacgaaagta cttaagtccc aaagctgtcg cacagcttag tccgcgcattt
     1261 gaaagtatt cttttatcacc tcagcagaag tctaaacgaa gacttttgc agagcaggac
```

```
1321 agcggactcg agttgacttt aacaaatgaa gctgaagatg tttcttctga ggtggaggta
1381 ccggctttag attctcagcc ggtagctggg gaacaatcag gggacataga catacatttt
1441 acagcattat tgcgtgcaaa caataacaga gcaattttaa tggcaaaatt caaggaagcg
1501 tttgggggtag ggttttatga cttaacacgc caatttaaaa gttataaaac atgctgtaat
1561 gcttgggtta tatctgtgta tgcagttcat gatgatctgc ttgaaagctc aaaacagctg
1621 ttgcaacagc attgtgacta tgtgtggatt agacagacag cagcaatgtc attgtttttta
1681 ttgtgcttta aagtgggaaa aaaccgtggc acggtgcata agttaatgat gtctatgtta
1741 aatgtacatg aaaaacaaat attatctgag ccacctaaac tgagaaatac tgctgctgcg
1801 ttattttggt ataagggctg tatgggatct ggagggttta cttatgtcc atacccagat
1861 tggatagcac aacaaacaat attaggtcat caaaatgctg aagcaagtag ttttgatttg
1921 tctgaaatga ttcaatgggc atttgataac aaccacatgg acgaatcaga tattgcgtat
1981 caatatgcaa aattagcacc agaaaacagc aatgctgtgg catggcttgc acataataat
2041 caagctagat ttgttagaga atgtgcggca atggtgcgtt tttataaaaa aggtcaaatg
2101 aaagaaatga gcatgtctga gtggatttat gctagaatcc atgaggtgga cggagaagga
2161 cactggtcca ctattgctaa attttttaaga tatcaacaag taaattttat aatgtttttta
2221 gcagccctaa aagatttatt gcatgcagtg ccaaaacgaa attgtatact aatttatggt
2281 cctcctaata ctgcaaaatc agcctttact atgtcactaa taagagtgtt aaaaggcaga
2341 gtaatatcgt ttgtaaactc caaaagtcag ttttggttac aacctttgtc agagtgtaaa
2401 atagcactct tagatgacgt cactgatccg tgttggattt acatggatac atatttgcga
2461 aatggtttag atgggcatgt tgtgtctttg gattgcaaac ataaagcccc cattcaaact
2521 aaatttcctg cgcttttact tacatctaat ataaatgtac ataatgaggt taattataga
2581 tacttacaca gcagaataca aggatttgag tttccaaacc catttcctat gaaagcagac
2641 aatacacctc aatttgaact tactgaccaa agctggaaat ctttttttac aaggctttgg
2701 caacaattag aacTGAgtga ccacgaagag gagggcgaaa ATGgagaatc tcagcgaacg
                 E2 orf start ->                    E2 cds ->
2761 tttcaatgtt ctacaagatc agctaatgaa catttaTGAa tctgcagcag aaacccttga
                                             <- E1 end
2821 gtcacaaatt gaacactggc aaattttgcg aaaagaagct gtactactat attttgctag
2881 gcgaaagggt gttacgcgaa ttggatatca acctgttccc acgttagcag tgtctgaagc
2941 aaaagctaag gaagcgatag ggatggtact gcagctgcag tcattacaaa aatctgaata
3001 tggaactgag ccttggtcct tggttgacac cagtgcagag acgtatagaa gtgctccaga
3061 aaattatttt aaaaagggtc caatgcctat agaggttata tatgacaaag atgcagataa
3121 tgccaatttg tataccatgt ggaaatttgt gtattacgtg atgaggatg acaattggca
3181 taaaagtgaa agtggggtaa atcatactgg tttatatttt atgcaaggaa actttagaca
3241 ctattatgtg ttatttgctg atgatgcacg taaatatagt gcaactggac attgggaagT
3301 AAagttaat aaggaaactg tgtttactcc tgtcaccagc tcaacacccc ccgactcacc
E4 orf start->
NH₂ terminus unknown
3361 aggaggacaa agagacccaa acacctcctc caagacccccc accaccacca ctgactccgc
3421 gtccagactc tcgcccacag cctccagaga acagtcacaa caaaccaaca ccaaaggacg
3481 gaggtacgag agacgacctt ccagcaggac ccgacgacaa acccaaacgc gccagaaacg
3541 atcaaggtcc aaatccaagt cccggtcgcg gtcgcggtcg cggtctcttt cgtctaaccg
3601 gcgatcacga tccaaatcca gaagaaaggc ctccaccact agagggagag gtcgagggtc
3661 acccaccgcc accagtgacc aatcctccag gtcaccctca gccacttcct ccacaacctc
3721 cttgcgatca gagagggagca gcagggtcgg gcgcagcagg gggggggcgca gcagggtcgg
3781 gcgcagtagg gggagaggga aacgatccag agagtcacca tcccccacca acaccaaacg
3841 gtcacgaaga cagtcagggt cttctaggct ccatggcgtc tctgctgacg cagtgggaac
3901 atcagttcac acggttagtg gaagaaatac aggaagactt ggaagattac tggaagaggc
3961 tcttgatccc ccagTAAttt tagttagggg ggaacctaac acgctacgta gctttcgcaa
                   <- E4 end
4021 cagagctaag cacatgtatc gagggctttt tagctcattt agcactgcat ggtcatgggt
4081 ggctggagat ggaattgagc gtctaggcag gaccagaatg ctcattagtt ttgtctcctt
4141 taatcaacga aagcactttg atgatacagt aaggtatcct aaaggtgtgg accgatcgtt
4201 tggctcattt gatagcctT AAcatactaa ctgcttttt tgctactaac acaaTAActt
                        <- E2 end                       L2 orf start ->
4261 acctatatat attttttac tgcaATGgcg cgcgccaggc gtactaagcg agattctgct
                       L2 cds ->
4321 accaatatat acagaacctg caaacaagca ggtacctgtc ctcctgatgt aatcAATAAA
                                                               signal ->
4381 gtagaacaaa ctacaatagc tgacaaaatc ttgcaatatg gcagtgctgg cgtcttttttt
4441 gggggtttgg gcataagtac tggcaaggac acaggaggag caactggtta cgtgcctttg
4501 gggggagggtc cagtacggtc tggtggtact gcaacgtgga tcaggccttc gttggttcca
4561 gacaccattg gaccatcaga cataatacct gtggacaccc tgaatccagt agagcccact
4621 acctcctcta ttgttccact aacagaggct tcaggttctg acctgttacc tggagaagtg
```

```
4681 gaaaccattg cagaggtgca ccctacacct agtatacct caacagatac cccagtgacc
4741 acaacttcaa gtggtgctag tgctgtttta gaagttgccc cagaacctgt tcctccatca
4801 cgtgtaagag ttactcgcac acaatatcat aatccttcct ttcaaatatt aactgaatct
4861 actccaacac agggtgaaag ctccttggct gatcatattc tggtcacttc aggttctgga
4921 ggacaaacaa ttggtagttc tggcagtgat ttaatagaac ttcaagaatt tcctactcgt
4981 tattcttttg agatagaaga acctacacct ccacgccaaa gtagtactcc aattcaaaga
5041 cttagaactg catttaggcg aaggggagga ttaacaaaata ggcgtttagt acaacaagta
5101 gctgtagatg atcccatatt cttaactcag ccttcaaggt tagtttcttt tcagtttgat
5161 aatcctgcat ttgaagaaga agtaacacaa atatttgaac aagatttaga taattttcgg
5221 gagccaccta atagggattt tttggatgtg caaactttag gtaggccaca atattcagaa
5281 acaccatctg gttacatcag agttagtcgc ctaggccaaa ggcgaaccat tcgcactcgc
5341 tctggagcac agataggttc acaagttcat ttctatagag acttaagtac tatagactcg
5401 gaagatccta tagagctaca attgttaggt cagcattctg gtgacgcttc aatagttcaa
5461 ggtaatacag aaagcacatt tataaatatt aatattgatg aaaatccatt agctgaagat
5521 tatagtatta ctgctaactc agaagatttg cttttagatg aagcacagga agactttagt
5581 gggtcacagt tagtagttgg tgggcgccgt tctacttcca catatacagt tccccaattt
5641 gaaactacaa ggtctggatc atattataca caagacacta aaggctatta tgttgcatat
5701 cctgaagata gaagtactag taaggataTA Atttatccca tgcctgactt gcctgtggtt
                        L1 orf start ->
5761 attatacata catATGacac cagtggtgat ttttacttac atcccagcct tcgcaagcga
            L1 cds ->
5821 ttcaaacgaa aacgtaaata tttaTAAttt tcttttgcag atggcagtat ggcaagcagc
                        <- L2 end
5881 tagtggtaag gtataccttc caccatctac accagttgcc agggtacaaa gcacggatga
5941 atatgttcaa aggacaaata tctactatca tgcttatagt gaccgcctac tcactgttgg
6001 tcacccatat tttaatgttt ataatgttgc aggatcaaaa ttagaaattc caaaagtttc
6061 aggaaatcaa cacagggttt ttagattaaga actaccagac cctaactgct ttgcacttgc
6121 tgatatgtca gtgtataatc ctgataaaga aaggttagtg tggggctgca gaggaattga
6181 aataggtaga gggcaaccac taggtgtagg tagtgtagga catccattgt ttaacaaagt
6241 aggggataca gagaaaccta actcatataa gggcacttct actgatgata ggcaaaatgt
6301 atcatttgac cctaaacagc tacaaatgtt tattataggc tgtgctccct gtatagggga
6361 acactgggat aaaagcattac catgtgctga gcaagatatt cctcagggat cctgtcctcc
6421 tatagagtta attaactcag ttattgaaga tggagacatg gcagatattg gctatggcaa
6481 tttaaatttt aaggccttac aacaaaacag atctgatgtc agtttagata tagtaaatga
6541 aacttgtaaa tatccagatt tcttaaagat gcaaaatgat gtgtatggtg attcctgctt
6601 tttttatgca aggcgagagc aatgttatgc tagacatttt tttgttcgtg gaggcaagac
6661 aggtgatgac attccagcag gacaaatcga tgagggaagc atgaaaaata catactacat
6721 acctcctaac aatagtcagc aacaatatac taatttagga aatgccatgt atttcccaac
6781 tgttagtggc tcattagtat ccagtgatgc tcaattgttt aacaggccat tttggttgca
6841 gcgcgcacaa ggtcataaca atggcatatg ctggtttaat cagctatttg tcacagtagt
6901 agacaacacg cgtaacacta attttagtat atcagttaat tcagatggaa cagatgttgc
6961 taaaattgca gattataatt ctgcaaactt taaagaatac ttaagacatg tagaagaata
7021 tgaaatatct ttaattttac aattatgtaa aataccttttg aaagcagaag ttctggcaca
7081 aatcaatgca atgaattcta acatattgga agaatggcaa ttaggtttcg tgcctgcacc
7141 cgataatcct attcaggaca cttatagata tatagattct ttagctacta gatgccctga
7201 caaaaatcct cctaaggaaa aagtagatcc ttataaaaac ttcacttttt ggaatgtaga
7261 tttatcagaa cgcctctctt tagatttaga tcaatatgct cttggccgca agtttttatt
7321 tcaggctggt ttgcaacagg caACCGTAAA CGGTacaaaa actatatctt cacggggtctc
                        -> E2 bind
7381 cagcagagga actaaaagaa agcgtaaaaa tTAAatttgt tcgttttcgt tacaataaag
                            <- L1 end
7441 tcaactttttg cacagtattc aaggaatgtt tatttactat gactaactaa gaaacgaacc
7501 gcacccgata cataaaggtg agttatgtgc caaatcagat acagtctgag ccgcatcagg
7561 cacagcagct ggccagatct gatcttcgtt gttttaacac gctcggatta ggactctcgc
7621 caatggaatc ataatcttgc caatctcttt tggcactgca cttggcaaag gTAAggACCG
                                                    E6 orf start ->
                                                        -> E2 bind
7681 TTAAC
```

```
LOCUS       HPV20E6        985 bp ds-DNA             VRL        15-JAN-1994
DEFINITION  Human papillomavirus type 20 (HPV-20), E6 gene.
ACCESSION   D90261
KEYWORDS    E6 gene.
SOURCE      Human papillomavirus type 20 DNA.
REFERENCE   1  (sites)
  AUTHORS   Ranst,M.V., Kaplan,J.B. and Burk,R.D.
  TITLE     Phylogenetic classification of human papillomaviruses: correlation
            with clinical manifestations
  JOURNAL   J. Gen. Virol. 73, 2653-2660 (1992)
COMMENT     Submitted (10-DEC-1990) to DDBJ by: Tohru Kiyono
            Aichi Cancer Center, Research Institute, 1-1 Kanokoden, Chikusa-ku
            Nagoya 464, Japan, Phone:  052-762-6111 x838, Fax:  052-763-5233.

            HPV-20 is primarily associated with the benign flat warts of
            Epidermodysplasia Verruciformis (EV), a multifactorial disease, and
            infrequently with malignancies.  HPV-20 is considered to be part
            of the a$_2$ cluster based on phylogenetic analysis.  This cluster
            includes HPV-19, HPV-25, HPV-14 and HPV-21, in addition to HPV-20.
            Patients with EV tend to have depressed cell mediated immunity.
            In roughly one-third of EV-associated HPV infection, the sun-exposed
            flat wart-like or macular lesions transform into malignant squamous
            cell carcinomas. Benign wart scrapings tend to be multiply-infected,
            with as many as six different viral types.  However, in contrast,
            EV carcinomas tend to harbor only a few types, specifically HPV-5
            and HPV-8, and less frequently HPV-14, HPV-17, HPV-20 and HPV-47.
            These types are rarely detected in lesions afflicting the general
            population.  A key to host restriction of these viruses may be in
            part due to the unusual organization of the LCR in these viruses.
            The LCR of these viruses is short compared to the viruses in other
            groups and contains two EV-specific regulatory regions: M33 and M29,
            both shown to be involved in protein binding.

            Kiyono et al. (Virology 186: 628-39) analyzed the noncoding region
            and the E6 gene of several EV associated HPVs (HPV-5, HPV-8, HPV-14,
            HPV-19, HPV-20, HPV-21, HPV-25, and HPV-47).  They observed that in
            the URRs of all eight of the EV HPVs a conserved region of 29
            nucleotides described by Krubke et al. (J Gen Virol 68: 3091-103),
            M29, was present.  Another conserved region of 33 nucleotides, M33,
            also described by Krubke , was common to only five of those types
            studied (HPV-5, HPV-8, HPV-47, HPV-25, and HPV-19).  Kiyono et al.
            established two operational clusters based on amino acid similarity
            of the E6 genes.  The first cluster consisted of HPV-5, HPV-8 and
            HPV-47; the second consisted of HPV-14, HPV-20, HPV-21, and HPV-25.
            In addition to genomic analysis, Kiyono et al. also determined the
            transforming ability of the EV types relative to HPV1a.  The
            results indicated that HPV-47, HPV-5 and HPV-8 had a stronger
            transforming ability than HPV-14, HPV-21 and HPV25; while HPV-1a
            showed no transformation.  Thus, the operational clusters hold for
            both genomic similarity and transforming ability.

BASE COUNT      294 a    174 c    215 g    302 t
ORIGIN
        1 tttatttact ctgactaact aagataccaa ccgcacccga cacataaagg tgagttgtgt
       61 gccaaatgag gtgagttgtg agccagaaga gatcacagcc aagtcaggct tgagccagat
      121 cagatacact gcgtgccaga gttggctcaa acttcatcgt cccaacacgt tcggaacagg
      181 aggaaatgta aggctgccaa cgcttttggc tcttcttttt ggcacagcag aagACCGTTA
                                                               -> E2 bind
      241 ACGGTaagtt tttatttgta tcgggcgcgg tcatacatta ctcatttggt agttgttgtt
      301 gccagctacc atcaagcata gcatgttttt gcctgtaacg ttatcggcac agtgattaaT
                                                               signal ->
      361 ATATATATAT ATATATATAT ATATATATAT ATATATagat atatagatac atatagacag
      421 atatcataga gctaatgcag agagtgcagg ctcATGgcta cacctccttc ttcagaagac
                                          E6 cds ->
      481 agcgctgatg aaggaccatc taatattgga gaggcaaaat ttccaatctt agagccacca
      541 ttgcctgcaa caatctgtgg cctagcgaaa ctttttagaaa taccgctaga tgattgtttg
```

```
601 ataccttgta acttctgcgg taatttcctt acacatttag aagtttgtga gtttgatgag
661 aagaagctta ctttaatttg gaaagatcat ttggtttttg catgctgtcg tgtttgttgc
721 tcggcaacag cgacatatga gtttaatcaa ttttatgaga gtactgtttt aggcagagac
781 atagagcaag taacaggcaa atctgttttt gatatacatg tcaggtgcta cacctgtatg
841 aaatttttag actcaattga aaagctagac atctgtggca gaaagcgtcc attttattta
901 gtgagaggct cttggaaagg aatctgtagg ctgtgtaagc attttcaaTA Atgattggta
                                                           <- E6 end
961 aagaggtcac attgcaagat attgt
```
//

```
LOCUS       HPV21E6       982 bp ds-DNA              VRL        15-JAN-1994
DEFINITION  Human papillomavirus type 21 (HPV-21), E6 gene.
ACCESSION   D90263
KEYWORDS    E6 gene.
SOURCE      Human papillomavirus type 21 DNA.
REFERENCE   1  (sites)
  AUTHORS   Ranst,M.V., Kaplan,J.B. and Burk,R.D.
  TITLE     Phylogenetic classification of human papillomaviruses: correlation
            with clinical manifestations
  JOURNAL   J. Gen. Virol. 73, 2653-2660 (1992)
COMMENT     Submitted (10-DEC-1990) to DDBJ by: Tohru Kiyono
            Aichi Cancer Center, Research Institute, 1-1 Kanokoden, Chikusa-ku
            Nagoya 464, Japan, Phone:  052-762-6111 x838, Fax:  052-763-5233.

            HPV-21 has been associated with the flat wart-like lesions of EV, a
            multifactorial disease.  HPV-21 is considered to be part of the
            a$_2$ cluster based on phylogenetic analysis.  This cluster
            includes HPV-19, HPV-25, HPV-14, HPV-20 and HPV-21.  Patients with
            EV tend to have depressed cell mediated immunity.  In roughly
            one-third of EV-associated HPV infection, the sun-exposed flat
            wart-like or macular lesions transform into malignant squamous
            cell carcinomas. Benign wart scrapings tend to be multiply-infected,
            with as many as six different viral types.  However, in contrast,
            EV carcinomas tend to harbor only a few types, specifically HPV-5
            and HPV-8, and less frequently HPV-14, HPV-17, HPV-20 and HPV-47.
            These types are rarely detected in lesions afflicting the general
            population.  A key to host restriction of these viruses may be in
            part due to the unusual organization of the LCR in these viruses.
            The LCR of these viruses is short compared to the viruses in other
            groups and contains two EV-specific regulatory regions: M33 and M29,
            both shown to be involved in protein binding.

            Kiyono et al. (Virology 186: 628-39) analyzed the noncoding region
            and the E6 gene of several EV associated HPVs (HPV-5, HPV-8, HPV-14,
            HPV-19, HPV-20, HPV-21, HPV-25, and HPV-47).  They observed that in
            the URRs of all eight of the EV HPVs a conserved region of 29
            nucleotides described by Krubke et al. (J Gen Virol 68: 3091-103),
            M29, was present.  Another conserved region of 33 nucleotides, M33,
            also described by Krubke , was common to only five of those types
            studied (HPV-5, HPV-8, HPV-47, HPV-25, and HPV-19).  Kiyono et al.
            established two operational clusters based on amino acid similarity
            of the E6 genes.  The first cluster consisted of HPV-5, HPV-8 and
            HPV-47; the second consisted of HPV-14, HPV-20, HPV-21, and HPV-25.
            In addition to genomic analysis, Kiyono et al. also determined the
            transforming ability of the EV types relative to HPV1a.  The
            results indicated that HPV-47, HPV-5 and HPV-8 had a stronger
            transforming ability than HPV-14, HPV-21 and HPV25; while HPV-1a
            showed no transformation.  Thus, the operational clusters hold for
            both genomic similarity and transforming ability.

BASE COUNT     300 a    158 c    209 g    315 t
ORIGIN
        1 tttatttact ctgactaagc aaataccaac cgcgcccgat acataaaggt gagttgtgag
       61 ccaaatgagg tgagttgtaa gccaaaagag gtcagagcca agtctgttct gagccagatc
      121 agatactacg cgcgccagag ttggatcaca tctcgttgtt ctaacacgct aaggactcaa
      181 ggaaatgtaa gtctgccaat cgattttggc tcgtgttttg gcagaagtga ggACCGTTAA
                                                                  -> E2 bind
      241 CGGTaagtta tgcACCGGGT GCGGTcgaat cattactcat ttgatagttg ttgttgccag
                   -> E2 bind
      301 ccaccattta ggacagcatg tttttgcctg taacgttatc ggcacatact cacaccaTAT
                                                                  signal ->
      361 ATATATATAT ATATATATAT ATATATATAT ATATATAAAT AAATATATAT ATATATATAc
      421 tagggagatg ctttagtact cATGgctgac tcttcaacag acagtgctga cgaaggtcct
                                             E6 cds ->
      481 tctcctaagc gtagacattt agaagaagaa aatacatcta gcttttttaga gccaccatta
      541 ccagctacaa ttcgtgacct agccaatctt ttagagatac cattggatga ttgtttagtg
```

HPV21E6

```
601 ccttgtaact tttgcggtaa ttttcttact catttagaag tttgtgagtt tgatgagaaa
661 aagcttagtt tactttggaa agatcattgt gtgtttgcct gttgtcgtgt ttgttgtgca
721 gcaacagcga catatgaata taatgaattt tatgaatcta ctgttgtagg tagagatata
781 gaagaaataa caggcaaatc tatttttgat attgatgtca ggtgctacac ttgcatgaaa
841 tttttagact caatagaaaa gctagatatt tgtggtagga agcatttttt tcataaagtg
901 agaggcactt ggaaaggaat ctgtaggctg tgtaagcatt ttcaaTAAtg attggtaaag
                                                        <- E6 end
961 aggtcacatt gcaagatatt gt
```

```
LOCUS       HPV25          7713 bp ds-DNA           VRL         04-OCT-1993
DEFINITION  Human papillomavirus type 25 (HPV-25), complete genome.
ACCESSION   X74471
SOURCE      Human papillomavirus type 25 DNA.
REFERENCE   1  (bases 1 to 7713)
  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 7713)
  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-25 was isolated by Gassenmaier et al. in 1984 (J Virol 52:
            1019-23) and subsequently sequenced by Dr. H. Delius.  It has
            been associated with the benign macular lesions of EV, a
            multifactorial disease.  HPV-25 is considered to be part of
            the a$_2$ cluster based on phylogenetic analysis.  This cluster
            includes HPV-14, HPV-19, HPV-20 and HPV-21, in addition
            to HPV-25.  Patients with EV tend to have depressed cell mediated
            immunity.  In roughly one-third of EV-associated HPV infection,
            the sun-exposed flat wart-like or macular lesions transform into
            malignant squamous cell carcinomas. Benign wart scrapings tend to
            be multiply-infected, with as many as six different viral types.
            However, in contrast, EV carcinomas tend to harbor only a few
            types, specifically HPV-5 and HPV-8, and less frequently HPV-14,
            HPV-17, HPV-20 and HPV-47. These types are rarely detected in
            lesions afflicting the general population.   A key to host
            restriction of these viruses may be in part due to the unusual
            organization of the LCR in these viruses.  The LCR of these viruses
            is short compared to the viruses in other groups and contains two
            EV-specific regulatory regions: M33 and M29, both shown to be
            involved in protein binding.

BASE COUNT     2357 a   1529 c   1699 g   2128 t
ORIGIN      199 bp downstream from beginning of E6 cds
        1 TAACGGTaag tctattgata cgggcgcggt taaattatta ctcattcgta cttgttgctg
          E2 bind <-
       61 ccaacaatca gcattagtaa cttgtttctg cctgtatcgt tatcgacaca ggtgtgttat
      121 acatatatat atatatatat atatatatat atatatatat attatatata tacacgtaga
      181 cactgcagca ttaggacttA TGgcaactgc aaatgctgaa cagagcatag gaccaccaga
                     E6 cds ->
      241 gcaagcgcag gttatacagc caccattgcc agcaacaatt actgatctag cagctttatt
      301 ggaaattcca ttagatgatt gcttagtacc ttgcaacttc tgtggcaact ttctaacata
      361 tttagagatc tgtgagtttg atgagaaaag acttagtttg atttggaaag aatatcttgt
      421 gtatgcctgc tgtcgctgtt gttgcacagc aactgccaca tttgaattta atgaatttta
      481 cgaaagcact gtaacaggta gggaaattga agacgttaca ggtaaatcaa tttttgacat
      541 agatgttaga tgtcaaacgt gcatgaaata tcttgatgca attgaaaagc ttgatatttg
      601 tggcagaaga cgtcctttcc atctagTAAg aggctcttgg aaaggaatct gtaggctgtg
                       E7 orf start ->
      661 taagcatttc tataATGatt ggTAAggagg tcacattgca agatttttaca ttagagttaa
                E7 cds ->         <- E6 end
      721 gtgaattgca gcctgaggta caaccagttg acctgttttg tgaagaggag ttgccggctg
      781 agcatcagga aacagaggag gagcctgcta ttgacagaac tccatacaaa gttgttgcgc
      841 cttgtggttg ctgcgaggtc aagcttcgca tctttgtgaa agccactgat tttggtatta
      901 gaacactaca aaaccttcTA Attgaagaac tgcagctgtt gtgtccggag tgtcgcggga
                 E1 orf start ->
      961 actgcaaacA TGgcggatcc TAAaggtagt acatctaaag aagggtttaa tgattggtgt
            E1 cds ->              <- E7 end
     1021 attttggaag ctgaatgtag tgatatagac aatgatttgg aacaattatt tgatcaagat
     1081 acagactcag atatttcaga cttattagat gagaatgacg tggaacaggg caattctcgg
     1141 gaactatttc atctacaaga gtgtcaggaa agcgaggagc aattgcaaaa actaaaacga
     1201 aagtacttaa gtccaaaagc tgtcgcacag cttagtccgc gattcgaaag tatttctttg
```

```
1261 tcacctcagc agaagtctaa acgaaggctc tttgcagagc aggacagcgg gctcgaattg
1321 actttaacaa atgaagctga agatgtttct cctgaggtgg aggtaccggc tttaaactct
1381 cagccggtag ctgagggaca atcaggggac atagacataa gttatacagc attattgcgt
1441 gccagcaata ataaagcaat attaatggca aaatttaaag aggctttttgg ggtagggttt
1501 aatgatctga cacgtcaatt taagagttac aaaacctgtt gtaatgcttg ggttatttct
1561 gtatatgcag tgcatgatga ccttatagaa agttcaaagc agcttttgca acagcattgt
1621 gactatgtgt ggatccgtgg gataggagcc atgtcattgt ttttagtttg ttttaaggcg
1681 ggaaaaaatc gtggtactgt tcataaattg atgacaacta tgttaaatgt gcatgaaaag
1741 caaatattat cggaaccacc aaaattaaga aatgttgctg cagcactgtt ttggtataaa
1801 ggatcaatgg gttccggagt atttacatat ggctcatatc cagattggat agcccaccaa
1861 acaatattgg gccatcaaag cgctgaagct agtacatttg atctatcgga catggttcaa
1921 tgggcatttg ataacaatta tttagacgaa gcagatattg catatcaata tgctaaattg
1981 gcgccagaca atagcaatgc tgttgcatgg ctcgcacata taatcaagc caaatttgtg
2041 cgagaatgtg catccatggt gagattttat aaaaaagggtc agatgaaaga aatgagtatg
2101 tctgaatgga tttatactaa aattcatgaa gtggaaggag agggtcaatg gtccaccatt
2161 gtacaatttt taaggtatca gcaagtcaac tttataatgt ttttagctgc cttaaaaagat
2221 ttactgcact ctgtccgaa acgaaattgt atactttttt atggacccc caatacgggg
2281 aaatcagctt ttactatgtc attgataaaa gtgttaaagg gtagggtatt atcattctgt
2341 aattccaaaa gtcaattttg gttacaaccg ttgtcagaat gtaaaatagc tttgctagat
2401 gatgtcacag acccctgttg ggtgtacatg gacacatatc tgagaaatgg cttagatgga
2461 cattatgtgt cattagattg taaacacaag gcaccaatgc agacaaaatt tcctgcacta
2521 ttgcttacat ccaatataaa tgtgcacaat gaagtgaatt atagatatct gcacagtaga
2581 ataaaaggat ttgagtttcc aaatcctttt ccaatgaaag cagacaaatac tccccagttt
2641 gatttaactg accaaagctg gaaatctttt tttacaaggc tttggcatca attagaccTG
2701 Agtgaccaag aagacgaggg cgaaaATGga gaatctcagc gagcgtttca atgttctaca
     E2 orf ->                E2 cds ->
     start
2761 agatcagcta atgaacattt aTGAaactgc agcacaaacc cttgaggcac aaattgagca
                          <- E1 end
2821 ttggcagatt ttgcgaagag aagctgtgct actatatttt gctaggcaaa agggtgttac
2881 acggcttgga tatcaacctg tacctgcctt aatggtgtct gaagcaaaag ctaaggaagc
2941 tatagggatg gtgctgcaac tgcagtcatt acaaaagtct gaatttggaa aagagccatg
3001 gtcattggtt gacaccagta cagagactta taaaagtcca ccagaaaacc atttcaaaaa
3061 aggcccaatg cctatagagg tcatttatga caaagatgca gataatgcca atgcttatac
3121 catgtggaga tatatttatt atgtggatga tgatgacaaa tggcataaaa gtgcaagcgg
3181 ggtgaaccac acaggcatat atttttatgca cggaagcttt agacactatt atgtgttatt
3241 tgctgatgat gctcgtagat atagcaatac tggacattgg gaagTAAaag ttaataagga
                                              E4 orf start ->
                                              NH2 terminus unknown
3301 cactgtgttt actcctgtca ccagctccac cccccccgag tcaccaggag acaagcaga
3361 ctcaaacacc tcctccaaga cccccaccac tgacaccgcg tccagactct cgcccacagg
3421 ctccggagaa cggtcacaac aaaaccagcac caagggacgg aggtacgaac ggaggccctc
3481 cagcaggaca cgacgacagc aagcccaagc gcgccagagg cgatcaaggt ccaagtcccg
3541 gtcccggtcc cggtcccagt cccgctcccg tatccgatcg aggtcgcggt cgcggtcgcg
3601 gtctgaatct cagtcgtcta agcggcgatc aagatccaga tccagaagaa aaacctcagc
3661 caccagaggg agaggtccag ggtcacccac aaccaccacc agtgaccgag ccgcaaggtc
3721 accttccacc acctcctctg ccacctccca acggtcacaa cgatcgcgat ccagggcagg
3781 gagcagcagg gggggccgcg gcaggggcgg gagacgacga cacagattgt ctgaatcccc
3841 cacctccaag cgatcacgaa gagagtcagg gtctgttagg ctccatggcg tctctgctga
3901 cgcagtggga acatcagttc acacagttag ttcaagacat acaggaagac ttggaagatt
3961 attggaagag gctcttgatc ccccagTAAt tttagttaga ggggatgcta cacgcttcg
                        <- E4 end
4021 aagctttcgc aacagggcaa agcatatgta tactgggcta tttagctcat ttagtacggc
4081 ctggtcgtgg gtggctggag atggcattga gcgtctaggc aggtccagaa tgctcattag
4141 ctttattttcc aacagtcaga gaaaacattt tgatgatgct gTGAgatatc cgaagggagt
                                              L2 orf start ->
4201 tgaccggtca tttggatcat ttgatagcct tTAAcctact aacattaggc tttgctacta
                                              <- E2 end
4261 acacactaac atatcttaac catttatttt tgtttttata tttttatgct ATGgcgcgcg
                                                            L2 cds ->
4321 ccaggcgggt taagcgagac tctgccacca atatatacag aacctgcaaa caagcaggca
4381 cttgtcctcc tgatgtgcta AATAAAgtgg aaaatactac aatagctgac aaaaattttac
                        signal ->
4441 aatatggcag tgctggtgtg tttttttcgggg gtttgggcat aagcactgga aaaggcacag
```

```
4501 gaggaacaac agggtatgta cctttgggag aaggtccgat acgtgttggt ggaACCCCCA
                                                                    -> E2 bind
4561 CGGTTattag accttcttta gtcccagaca ctattggacc atctgacata atacctgttg
4621 acactcttaa tccagtggaa cccacatcct cgtctattgt cccactcaca gagtcttcag
4681 gtcctgacct tttgcctggt gaggtggaga caatcgcaga aatacatcct gggcctgttg
4741 taccctccac tgacaccccg gtgacaacaa cttctagagg tgccagcgca gttttagagg
4801 ttgcaccaga gcccaccccg ccatcgcgcg tcagagttag tggcacacaa tatcataatc
4861 catcatttca ggttataact gaatccacac ctcaag tgaaagttca ctggctgatc
4921 acattttggt tacttctggt tctggagggc aaacaattgg tggcactgcc agtgacctaa
4981 ttgaactaca agagtttccc actcgctatt catttgaaat agatgaaccc acacctccaa
5041 gacaaagtag tacacctctt caaaggatta ggactgcatt aagacgtaga ggaggattaa
5101 caaatagacg attggtgcaa caggtacctg tagaagaccc tttattttg tctcagccat
5161 cacggctagt acggtttcag tttgacaatc cagtatttga ggacgaagtt acacaaattt
5221 ttgagcagga tttaaatgat tttcaggagc ctcctgacag ggattttta gatattcgct
5281 cattaggaag gccacaatat tctgaaactc cggctggcta tgtacgggtc agtcgccttg
5341 gtcaaagacg taccattcgc actcgatctg gggctcaaat aggctcacaa gttcattttt
5401 atagagactt aagcagtata aatactgaag atcctattga gctacagttg ttaggtcagc
5461 attcgggtga tgctacaata gttcaaggcc tcacagaaag tacctttgta gatgttaatg
5521 tagatgaaaa tccattagct gaagatttta gtattctgc acactctgat gatttacttt
5581 tggatgaagc taatgaagat tttagtggtt cccagttggt tgtgggtggt cggcgctcca
5641 cttctactta cactgtgcct cgtgttgaaa ctacacgtc tgcatcatat tatacccaag
5701 atattcaggg atactatgtg tcctatcctg aggatagaga tactagtaag gacattattt
5761 accctatgcc tgacctcccg gtggtcatta tacacacata tgacaccagt ggtgattttt
5821 atttgcatcc cagtctgact acaaggcgca gacgcaaaag aaaatattta TAAtttttc
                                                                    <- L2 end
                                              L1 orf start ->
5881 ttttacagAT Ggcagtttgg caagcagcta gtggtaaagt gtaccttcca ccatctacac
        L1 cds ->
5941 ctgttgccag ggtacaaagc acggatgaat atgtgcaaag aactaacatc tattatcatg
6001 cctatagtga ccgcctatta actgttggtc acccatattt taatgtgtac aacgtccaag
6061 gctctaaatt gcaaattcca aaagtgtcag gaaatcaaca cagagttttt aggttaaaat
6121 taccagatcc caatcgtttt gctcttgcag acatgtcagt ttacaatcct gacaaagaaa
6181 ggctggtttg ggcctgcaga ggtattgaaa taggacgtgg gcaacctta ggtgtgggaa
6241 gtgtgggtca cccattgttt aacaaggttg gcgacacaga aaatccaaat tcttataaag
6301 ctagttctac agatgacagg caaaatgtat catttgaccc caaacaacta cagatgttta
6361 ttataggctg tgctccatgt ataggggaac attgggataa agcattacct tgtgatgatg
6421 gcaatattca acaagggtca tgccctccaa tagaattaat taattctgtc attgaagatg
6481 gggatatggc agacattggc tatggtaatt taaattttaa agcattacag caaacagag
6541 ctgacgtaag tctggatata gttaacgaaa cctgtaaata tccagacttt cttaaaatgc
6601 aaaatgatgt gtatgggat tcctgctttt tttatgcacg gcgggaacaa tgttatgcca
6661 gacatttttt tgttagaggg ggcaaaacgg gtgacgatat tccagctgga caaattgatg
6721 aaggaagcat gaaaaatgca ttttacatac cacctaacag tagtcaggct caatataata
6781 atctaggtaa ctcaatgtat ttcccaacag tcagtggctc attggtatcc agtgatgctc
6841 aattattcaa taggccattt tggttacagc gagcacaagg tcataacaat ggtatatgct
6901 ggtttaatca gctatttgtc actgtggtag acaacacacg caacactaat ttcagcatat
6961 caatcaattc agatggaaca gatgtttcca aaatcactga ttataattct caaaaattta
7021 cagaatattt gagacattga gaagaatatg agttatcatt aatattacaa cttttgcaaag
7081 taccgttgaa ggcagaaata ttggctcaga ttaatgcaat gaattccaac attttagaag
7141 aatggcaatt aggattcgtt cctgcaccgg acaattctat tcaggatact tatcgctaca
7201 ttgattcttt agccacacgt tgtccagata aaaatcctcc taaggaaaaa gtagatcctt
7261 ataaaaattt acactttttgg gatgtagatt taacagaacg cctttctta gatttagatc
7321 aatattcact gggacgtaag tttttattc aggctggttt gcagcaaaca ACCGTAAACG
                                                                    -> E2 bind
7381 GTacaaaaac agtttcctcc cgaatatcta ctaggggaat aaaaagaaa cgtaaaaatT
7441 AGaaattacc gctttcgata caataaagtc aacttttgca cagtattcaa ggaatgttta
        <- L1 end
7501 tttgtgctga ctaactagga taccaaccgc acccgataca taaaaggtga gttatgtgcc
7561 aaaatcagata cagtctgtgc caaactcagg cagctgctcg ccagatgcgt atcgtcttta
7621 gtgtttttgac acgctcggat taggacgttc gccaatgaa tacTAAtatt gccaatcgct
                                              E6 orf start ->
7681 ttcggctctt tgtttggcag agctcaagAC CGT
                                       -> E2 bind
```

```
LOCUS       HPV47        7726 bp ds-DNA           VRL      15-SEP-1990
DEFINITION  Human papillomavirus type 47 (HPV-47), complete genome.
ACCESSION   M32305
SOURCE      Human papillomavirus type 47 DNA isolated from scrapes of a benign
            lesion of a patient suffering from EV and skin cancer.
REFERENCE   1  (bases 1 to 7726)
  AUTHORS   Kiyono,T., Adachi,A. and Ishibashi,M.
   TITLE    Genome organization and taxonomic position of human papillomavirus
            type 47 inferred from its DNA sequence
  JOURNAL   Virology 177, 401-405 (1990)
COMMENT     Draft entry and printed sequence for [1] kindly submitted by
            T.Kiyono, 23-FEB-1990, for release after publication.

            HPV-47 is primarily associated with the benign lesions of EV, a
            multifactorial disease, and also has been detected in
            cases of malignancy.  HPV-47 is considered to be part of
            the a$_1$ cluster based on phylogenetic analysis.  This cluster
            includes HPV-5, HPV-8, HPV-12, and HPV-47.  Patients with EV tend
            to have depressed cell mediated immunity.  In roughly one-third of
            EV-associated HPV infection, the sun-exposed flat wart-like or
            macular lesions transform into malignant squamous cell carcinomas.
            Benign wart scrapings tend to be multiply-infected, with as many as
            six different viral types.  However, in contrast, EV carcinomas
            tend to harbor only a few types, specifically HPV-5 and HPV-8, and
            less frequently HPV-14, HPV-17, HPV-20 and HPV-47. These types are
            rarely detected in lesions afflicting the general population.   A
            key to host restriction of these viruses may be in part due to the
            unusual organization of the LCR in these viruses.  The LCR of these
            viruses is short compared to the viruses in other groups and
            contains two EV-specific regulatory regions: M33 and M29, both
            shown to be involved in protein binding.

            HPV-47 has been isolated from the scraping of a benign lesion of a
            patient who had suffered from both EV and skin cancer.  The DNA was
            subcloned into vector pTZ18R and the complete sequence was
            determined.  Recently Kiyono et al. detected HPV-47 by PCR in small
            clusters of malignant cells in the lower dermis of the same patient.
            Also a splice donor/acceptor pair which if used can result in a
            E1/E4 fusion product is present in this genome.  Both HPV-5 and
            HPV-8 also have this ability.

BASE COUNT    2369 a   1517 c   1727 g   2113 t
ORIGIN      207 bp upstream from beginning of E6 cds
        1 AACGGTaagt ttgcattaat gtACCAGGTG CGGTacagat catttcacaa tggatattat
    E2 bind <-          E2 bind ->
       61 tgttgccaac taccatagtc ataatcaagt tcttgcctgt atcgttttcg taccttacct
      121 acagtatttt atattaaTAT ATAAataaat aaaTATATAA atgtgtattt atttctcagg
                    E6 orf start ->
                    signal ->          signal ->
      181 ctcagttctt tgcaattatt aagacaaATG gctcagaagg ctttggaaca gactacagtt
                             E6 cds ->
      241 aaagaggaaa agctagaact acctactact attagaggct tagctcaatt gttagacata
      301 cctttagtag attgtttgct accttgcaac ttttgtggca gatttcttga ctatttagaa
      361 gtttgtgaat ttgattataa aaagcttact ttaatttgga aagactacag tgtttatgcc
      421 tgctgccgtt tgtgctgctc agcaactgcc acatatgaat ttaatgtttt ttatcaacaa
      481 acagtgttag gtagagatat tgagctagct acaggccttt ccatttttga gattgacata
      541 aggtgtcata cctgcctgtc atttcttgac attattgaaa agttagatag ctgtggaaga
      601 ggacttccct ttcacaaagT AAgaaacgcc tggaagggtg tttgtaggca gtgtaagcat
                    E7 orf start ->
      661 ttttacaATG attggTAAag aggtcaccgt gcgagatatt gttctggagt taagtgaggt
            E7 cds ->          <- E6 cds end
      721 tcaacctgaa gtattaccag ttgacctgtt ttgcgacgag gaattaccaa atgaacaaca
      781 ggcggaggag gagctagaca tcgacagagt cgttttcaaa gtgattgcac cgtgcggttg
      841 cagctgctgc gaggtcaagc ttcgcatttt tgtgaacgca acaaaccgtg gcatcaggac
      901 atttcaggaa ctttTGActg gtgatctgca gctcctctgc ccagagtgcc gtgggaactg
                E1 orf start ->
```

```
 961 caaacATGgc ggattcTAAa gGTagtacat ctaaagaagg gtttggtgat tggtgtattt
     (E1/E4 fused orf) 5' sj /\
       E1 cds ->             <- E7 cds end
1021 tggaagctga ctgtagtgat gttgaggatg atttgggaca attatttgag agagatacag
1081 actcagatat ctcggacctg ttagacaatt gtgacctgga tcagggcaat tcacgggaac
1141 tatttcatca acaggagtgt aagcaaagcg aggagcaatt acaaaaacta aaacgaaagt
1201 atcttagtcc aaaagctgtc gcgcagctta gtccgcgtct tgagtcaatt tcattgtcac
1261 ctcagcagaa atccaagaga aggctctttg cagagcaaga cagcggactc gagttaacct
1321 ttaacaatga agctgaagat gttactcctg aggtggagGT accggctata gactctcggc
                                         5' sj /\
1381 cggatgatga tgaggggagga tcaggggatg tagatattca ttatacagca ttgttgcgtt
1441 ccagcaacca aaaggccaca ttactggcaa aattcaaaca agcgtttggg gtaggcttta
1501 atgaattgac aagacaattc aaaagctaca aaacctgctg taatcattgg gttgtatccg
1561 tatatgcagt ccatgatgat ctatttgaaa gctcaaagca gctggttcaa cagcattgtg
1621 actatatatg ggtccgtggg atagatgcaa tgtcattata tctattgtgt tttaaggcgg
1681 gaaaaaatcg tgggacagtt cataagctaa ttaccacaat gttaaatgtg catgagcaac
1741 agatattgtc tgagcctcca aagttaagaa atacagctgc tgcattattt tggtacaaag
1801 gatgtatggg acctggagtg ttcacccacg gtccttaccc tgaatggatt gcacaattaa
1861 ccattttggg ccataagagt gctgaggcaa gtgcgtttga tctgtcagtc atggttcaat
1921 gggcatttga taacaatctg tttgaggagg cagacattgc atacggatat gcaagactgg
1981 caccagagga tagcaatgca gttgcatggc ttgcacataa taaccaagct aaatatgtta
2041 gagaatgtgc tatgatggtt cgatactaca aaaaggggca aatgagagat atgagcatgt
2101 ctgagtggat atatacaagg atacatgaag tagagggaga aggacagtgg tctagcattg
2161 ttaaattttt aagatatcaa gaaataaaat ttatttcatt tttggctgct ttaaaagatt
2221 tattcattc agtacctaaa cgcaattgta tttttattcca tggccctcca aatacaggaa
2281 agtcatcgtt tggaatgtcc ttaataaaag ttctaagggg gagagtatta tcatttgtaa
2341 actccaaaag tcagtttttgg ttgcagcctc ttggagaatg taaaatagca ttattagatg
2401 atgttacaga tccatgttgg gtgtatatgg atcaatattt aagaaatggg ttagatgggc
2461 atttgtgtc tttggattgt aaatatagag cacccatgca aacaaagttt ccaccttaa
2521 tacttacatc taatattaat gtacatgcag agaccaatta tagatacca catagtagaa
2581 ttaagggttt tgaatttaaa aatccatttc ctatgaaagc agataataca cctcaatttg
2641 agttaactga ccaaagctgg aaatctttt ttacaaGGct ttggacacac ttagaccTGA
                                     /\ 3' sj    E2 orf start ->
2701 gtgaccaaga agacgagggc gaacATGgag aatctcagcg agcgtttcaa tgctctgcaa
                      E2 cds ->
2761 gaacagctaa tgaacattta TGAagctgca gaacagacat taaaggcaca aattttacat
                    E1 end <-
2821 tggcagacat tgcgaaaaga agctgtgaca ctctactttg ctaggcagaa aggcataaat
2881 aggttgggat accaaccagt gcctgcatta gcaatatctg aggcaagggc caaagaggct
2941 atatatatg tgttcagtt agagtcgcta caaaaatcag cgtttgcttt ggagccttgg
3001 accttagtgg acactagtac agagactttt aagagtgctc cagaaaatca ttttaaaaag
```

```
3061 gggcctgtac ctgtggaggT GAtatATGac aaagatgaag caaatgctaa tttgtatact
                              E4 cds ->
                 E4 orf start ->
3121 atgtggacat ttgtgtatta catggattca gatgatgtgt ggcataagac aacaagtggg
3181 gtcaatcaaa ctggcattta ctacctatat ggaacattta aacactatta tgtgttattt
3241 gctgatgatg caaagagata tagtgctact ggagaatggg aagttaaagt taataaggaa
3301 actgtgttta ctcctgtcac tAGctccaca ccaccagggt caccaggagg acaaacagac
               (E1/E4 fused orf) /\ 3' sj
3361 ccagacacct cctccaagac ccccaccacc accacagccg ccactgacac ctcgcccaga
3421 cgccaatcca tcaataaaca gtcacaacaa accgaaacca aacgaagagg gtacggacgg
3481 agaccatcaa gcagaacaag gcgaccgcaa acgcaccaaa ggcgatccag atccagatcc
3541 cggtcgcggt ccagttctca aacccactct tccaccacca ccaccaccac cacctacagg
3601 tccaggtcta cgtcgctcaa caagactcgt gctcgttcca ggtcaaggtc cacctccaga
3661 tctaccagca ccaccagtag aaggggaggt agagggtcat ccacaaggca agatcgcga
3721 tcaccctcca cctacacctc aaaacggtca cgggaaggaa acacaagggg cagagggagg
3781 gggagacaag ggagagcagg gagcagtggg gggagagagc agcgacggag aaggagatca
3841 ttctcaacct cccctgactc ctccaaacga gtcagacggg agtctcctaa ataccgtggc
3901 gtgtctccta gcgaggtggg aaagcaactt cgatcagttg gtgcaaaaca ttcagggcga
3961 cttggaaggt tattggagga agctagggac cccccagTAA ttcttgtgcg aggggacgca
                                                  <- E4 end
4021 aacacattaa aatgctttcg caacagagca aggaacaaat atagagggct ttttagatca
4081 ttcagcacta cattttcctg ggtagctgga gatagcattg agcgtctagg caggtccaga
4141 atgctcatta gcttttcctg cctcactcag agaagggatt ttgatgatgc tgtcaaatat
4201 ccaaaaggag tcgagtggtc atatggtagt cttgatagcc tttaacaagc attaacgctg
4261 ctttgctact aactgctatt aacaaccaca gctttttttt tacgttttt tattttacTG
4321 Attttgtact gcaATGgcgc gtgctagaag ggtcaaacgt gactctgtaa cacatatata
L2 orf ->    L2 cds ->
  start
4381 tcagacctgc aaacaggcag gcacttgccc ctcggacgtt gttAATAAAg ttgagcaaac
                                                signal ->
4441 aacagttgct gacaatattt tgaaatatgg cagtgctggt gtctttttg gaggccttgg
4501 cataggaaca ggccgaggga ctgggggtgc tactgggtac gtgccacttg gggaaggtcc
4561 tggtgtccgt gtgggaggaa ccccaacggt tgtaaggcct tctcttgttc ctgaagcaat
4621 tggaccagtt gatattttac ccattgacac aatcgcacct gtcgagccta ctgcttcatc
4681 tttagtccca ttaacagagt cgtctggtgc tgatttactt cccggtgaag ttgaaactat
4741 agccgaaata catcctattc ctgaaggtcc gacaatcgac tcccctgtag tcaccacaac
4801 gacaggttcc agtgctgttc tggaagtggc tccagaacct gtacccccta cacgtgttag
4861 aattgctaga acacaatatc ataatccctc ttttcagata ctcactgaat caacacctgc
4921 gcagggcgag agttctcttg ctgaccatat tttggtcacc tcaggtctg gtggacaaag
4981 gataggcggt gatataacag acgaaattga acttactgag tttccaagca gatatacatt
5041 tgaaatagaa gaacccaccc ctccacgaaa aagtagcaca ccattacaaa ctgtagcctc
5101 tgcagtaagg cgacggggct ctctcattaac aaatagaaga ttggtacaac aagtagctgt
5161 agacaatcct ttattttaa gtcaaccttc taagatggta agattctcat ttgacaatcc
5221 agcttttgaa gaagaggtta ccaatatttt tgaacaggat gttaacagct ttgaagaacc
5281 tccagacagg gattttcttg atattaaaca attgggccgt cctcaatatt ctacaacacc
5341 agcaggttat attagggtaa gcagactagg aactcgaggc accattcgca ctcgttctgg
5401 tgcacaaata ggttctcagg tacactttta gagatttta agttctataa atactgagga
5461 tccaatagaa ctacagtttt tagggcagca ttctggagat gctactattg ttccaaggtcc
5521 tgtagaaagc acatttatag atatggacat tgctgaaaac cctttatctg aaacaataga
5581 tgcttcatct aatgatttac ttttggatga gactgtggag gattttagtg ggtcccaatt
5641 agtaattgga aatcgaagga gtacaacatc atatactgtt cccagatttg agactactag
5701 aagtagttcc tattatgttc aagacacaga tggttattat gttgcttacc cagagtcacg
5761 ggacactatt gatattattt accctacacc tgaattacct gtagttgtca ttcacaccca
5821 tgacaattct ggagactttt acttacatcc tagtcttaga aggcgtaagc gtaaaagaaa
```

```
5881 atatttgTGA tttgcattgc agATGgcagt gtggcactcg gctaacggta aagtatacct
        L1 orf start ->    L1 cds ->
                <- L2 end
5941 tcctccatca acaccagtgg ccagggttca aagcacggat gaatacatac aaaggactaa
6001 tatctattat catgcaaata ctgaccgcct tttaacagta ggacatccat atttcaatgt
6061 atacaataat aatggaacta cattagaggt tccaaaagta tcaggtaatc agcatagggt
6121 gtttcgctta aaattgccag atcctaatag atttgctcta gcggacatgt ctgtatacaa
6181 ccctgacaaa gaacgcttgg tgtgggcctg caggggtcta gaaattggaa ggggtcaacc
6241 tttaggtgtt ggcagtactg gtcacccata tttaataag gtaaaagata cagaaaacag
6301 taattcctat atcacaaact caaaagatga cagacaagac acctctttg atcctaaaca
6361 aatacagatg tttattgtgg gctgcactcc atgtattggc gaacactggg ataaggcaga
6421 gccttgtggg gaacagcaaa ctggtctttg tcctcctatt gaattaaaaa acacatacat
6481 tcaggatggc gacatggcag acattggttt tggcaacatt aatttcaagg ccttacaaca
6541 cagtaggtct gatgttagtc ttgacattgt aaatgaaact tgcaagtacc cggattttct
6601 caaaatgcaa aatgatgttt atggggatgc ttgcttttt tatgctcgta gagagcaatg
6661 ttatgccaga cattttttg ttagaggggg aaaaacaggt gatgacatac caggagcaca
6721 ggttggcaat ggtaatatga aaaatcaatt ttacattcct ggtgctacgg tcaggctca
6781 gagcactata ggtaatgcca tgtatttccc aactgtcagt ggctcactag tctctagtga
6841 tgctcaactg tttaacaggc cattctggct ccaaagggct cagggtcata taatggcat
6901 tctgtgggct aatcaaatgt ttgtcacagt tgtagacaac acaagaaata caaatttcag
6961 catctctgtt tactctcagg caggggacat aaaggatata caggattata atgcagacaa
7021 ttttagagag tatcaaagac atgtggagga atatgaaatt tctgtaatat tacaattgtg
7081 caaagttcct ttaaaagcag aagttttagc acaaattaat gccatgaatt cgtctctttt
7141 agaggaatgg cagttaggat ttgtgcctac tccagacaac cctattcagg atacatatag
7201 atatctagaa tctttggcca ctaggtgtcc tgaaaagtct cctccaaaag agaaggttga
7261 ccccctacaaa ggtttaaaact tttgggatgt cgatatgaca gagcgccttt ccctggattt
7321 agatcaatat tcattaggta gaaagttctt attccaggct ggattacagc agacgACCGT
                                                            E2 bind ->
7381 AAACGGTaca aaaacaactc cttacagggg gtccatcaga ggaacaaagc gcaaacgaaa
7441 aaatTGAaga tgACCGTTTT CGGTacagat tgtttaactt ttacacagta ttcaaggaat
            <- L1 end
                    -> E2 bind
7501 gtctgtttac tgtgactaag tgtaactctg ccaaagaaac aACCGCACCC GGTacacgta
                                                E2 bind ->
7561 ttcagcttgt tgccaaaaca gataagcttg gcagtcagaa cacaccgtgt tcgtcgcaac
7621 acgctcggat taggtcttct gccaaaagaa atttaatctt gttatcgttt ttggcgatca
7681 catttggcac cgcgggcagc tgttttggca ctacaagaca ACCGTT
                                                E2 bind ->
```

```
LOCUS       HPV49        7560 bp ds-DNA           VRL       04-OCT-1993
DEFINITION  Human papillomavirus type 49 (HPV-49), complete genome.
ACCESSION   X74480
SOURCE      Human papillomavirus type 49 DNA.
REFERENCE   1  (bases 1 to 7560)
  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 7560)
  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-49 was originally isolated from pooled flat warts of a Polish
            renal transplant patient (Favre et al.  J Virol 63: 4909 (1989)
            and subsequently seqeunced by Dr. H. Delius.  Favre et al. screened
            benign cutaneous lesions from 134 patients, including 51
            immunosuppressed patients and 35 epidermodysplasia verruciformis
            patients, premalignant cutaneous lesions from 64 patients, and
            invasive skin carcinomas from 48 patients, for HPV-49 DNA.  Despite
            its similarity to other EV-related papillomaviruses, HPV-49 was
            detected only in the flat warts of two other Polish renal
            transplant patients, and in none of the EV patients.  HPV-49 forms
            a remote branch off of the b cluster of the EV- associated HPV
            types.
BASE COUNT     2366 a   1436 c   1672 g   2086 t
ORIGIN      199 bp upstream from beginning of E6 cds
        1 CCACATTCGT Tccagctaca ttttggcgcc aactctttgg cagcaacacc agaacgataa
          E2 bind <-
       61 cggtaagttt caatcgggcg cggtcacatt atacttagtc atctcttgtg gttgttaaca
      121 acaatctTGA aacagatata catgtaaccg cttgcgtgct gtactttctt tattcttgga
        E6 orf start ->
      181 aagaatacag acaggacacA TGgctagACC TGTTAAGGTa tgtgagctag cccaccactt
                      E6 cds ->        -> E2 bind
      241 aaatatacct atttgggaag tttttgcttcc ttgtaatttt tgcacggggt ttctaacata
      301 tcaggagttg ttagaatttg actataaaga ctttaatttg ctgtggaaag acggatttgt
      361 ctttggttgt tgtgcagctt gtgcctatag atcagcatat cacgagttta ctaattatca
      421 ccaagaaatt gtcgtaggca tcgaaataga aggacgagca gcggctaata ttgctgagat
      481 agtagtcaga tgtctcattt gccttaagag gctagatttg ttggaaaagc ttgatatttg
      541 tgcacagcac agagagtttc acagagttag aaataggtgg aaaggggtgt gTAGacattg
                                                        E7 orf start ->
      601 cagagttata gaATGAttgg gaaagaagtt acaataccag atataatact acaagaagag
              E7 cds -><- E6 end
      661 tttggccagc ccattgacct gcaatgctac gagaatctaa cagctgaagc gccagctgaa
      721 caagagttgg aggcagagga ggagcttatc caaggcatcc cttacaaagt tattgctact
      781 tgtggcggcg gatgcggtgc cagactgcga gtcttcgtgt TAGccactga cgctgctatt
                                                E1 orf start ->
      841 agaagtttcc aagaactgct tctggaggaa ctgcaattct tgtgtcctca gtgtcgtgaa
      901 gaaattcgga ATGgcggacg aTAAaggtac tgatcccaaa gaagggtgta gcgagtggtt
            E1 cds ->          <- E7 end
      961 tatagataat gaagcagact gtagtgattt agaaaatgat ttggaacaat atttgatga
     1021 aagcccaaag tccaatattt caaatttgtt aaatgatgag gaggatgtgg agcagggaaa
     1081 ttcgcgagat ctgcttcgcc agcaggaatt tgaggagagc gcggagcaag tacaaaagtt
     1141 aaaacgaaag tatttcagtc ctaaagcagt tcaacaactt agcccacggt tgcagtctat
     1201 gtcaatatct ccgcgacaaa agtctaaacg aaggctattt gaggaggaca gcgggctgga
     1261 attatcgggg ctcgaacagt ctttgactaa tgaaattgaa gatactcctg cggagctgga
     1321 ggtaccggcg gcaacgccgg cagagcaggg tggtcaggga gagggcaatt tgcattataa
     1381 agagttaatg cgatgcaata atagtcgtgc aaaattatta agtaaagtca aggaatattt
     1441 tggtgtgggt ttttatgagt tagctagaca gtataaaagt gataaaacat gctgtaaaga
     1501 ttgggtaatt gcagcctatg gcgtgcgaga gagctggta gaaagtgcaa aacaattact
     1561 tttaaatcat tgttcctatg tgtggataaa tataatggg attatgactt tatatttact
     1621 gtgttttaat catgcaaaga gtagagaaac tgttggtaga ttgcttatgt caatactgga
     1681 tgtacaatta ttgcaattaa tttgtgaacc accaaaacta agaagtgtgg tgtcagcact
```

```
1741 atactggtac aaaggcagta tggactcatc tgtgtatgct catggagcct atcctgattg
1801 gattgtaaat cagaccatga taagtcatca ggcagcagca gatgctatgc aatttgacct
1861 ttctgaaatg atacaatggg cctatgatag cgatctcaca gatgaagctg acattgcata
1921 tctttatgct aaaatggcaa atagtgactc taatgcaaga gcttggttag cacataataa
1981 tcaggcaagg tacttaagag aatgtgctca aatggttaga cattacagac ggggagaaat
2041 gagggatatg agtatgtctg agtggataca tcacagaata caacaagtag aaggggaagg
2101 ccattggtct gaaatagtta agtttataag atttcaagaa ataaacttta taatatttct
2161 ggatgcattt aaacagttta tacatggcaa acctaaaaaa agctgtttat taatacatgg
2221 gccgccggac tgtggcaagt caatgtttgc tatgtcatta ttaaaagttt taaaaggcaa
2281 ggtaatttca tttgtaaatg caaaaagtca attttggctg tctccacttt cagaatgtaa
2341 aatagggctg ttggatgatg ctaccgatcc ttgttggcaa tatatagata catatttaag
2401 aaatggtctc gatggaaatg ttgtaagtgt ggattgcaaa cataaaaccc ctatgcaaat
2461 taggttccca ccattgttaa taacttcaaa ttataatatt aaagctaatg ataaatataa
2521 gttttttgtac agtagaattg caatatttga atttaaacat aagttcccat tcaaagagga
2581 tggtacccct gtatttcaac ttactgacca aagctggaaa tcttttttttg aaaggctttg
2641 gacacaatTA Gagctcagtg acccagaaga cgaggcagac aATGgaggca ctcaacgctc
         E2 orf start ->                         E2 cds ->
2701 gtttcaatgt actacaagag atgttaatgg acatttaTGA atcagggaaa gaggatcttg
                                          <- E1 end
2761 aaacacaaat agaacattgg aaactgttaa gacaggaaca agctttatta tttttttgcac
2821 gtaaacacag cataatgaga ctggggtatc aacccgtacc tccgatggca gtatctgaaa
2881 ccaaagccaa acaagctatt ggcatgatgc taactttgca aagcttgcaa aagtctcctt
2941 ttggaaaaga aaagtggact ttagtaaaca caagtcttga aacatacaat gcaccaccag
3001 cacagtgctt taaaaaaggt ccttataata tagaagttat atttgatgga gatcctgaaa
3061 atctaatggt atatactgct tggaaagaga tttattttgT AGactcagAT Gatatgtggc
                                                  E4 cds ->
                                        E4 orf start ->
3121 aaaaggtgca aggtgaggtg gattatgcag gtgcatatta taaggatgga actatcaaac
3181 agtattatgt taccttcgct gatgatgctg ttagatatgg gacatctgga caatatgaag
3241 tccgcattaa caacgaaact gtgtttgctc ctgttactag ctccacccca ccatccacgg
3301 ggctacgaga atcctccaac gccagccccg ttcacgacac cgtcgacgag acacccacca
3361 gcaccacagc aaccaccacc accttcagca ccaccacagc cacagccaca gccacaggag
3421 cacctgaact ctcatccaaa accggtacca ggaaaggaag gtacgggcga aaagactcta
3481 gtcctacagc agcctccaac tccaggaaag aggtctcgcg acgacgatcc aggtctagaa
3541 ccaggacccg cagacgggaa gcgagcacct caaggtccca aaaagccagc cgttccagat
3601 ccagatccga atccacttcc agaggatcca gaggtccgg aggatctgtc acaacctcca
3661 gagattccag ccccaagaga acccgcaggg gcagaggaag gggaggaga agtagaaggt
3721 cacccacccc cacctccacc agtaaacggg aaagaaggc cagccggtca aggggggag
3781 agcctgtttc tggaggggtt ggcatctcgc ctgacaaggt gggatcaaga gtacaaacag
3841 ttagtggacg acatcttgga cgacttggaa ggttactgga ggaggctagc gatcctccag
3901 TAAtactttt gcgaggagac ccaaatattt taaaatgtta cagatacaga gataagaagc
         <- E4 end
3961 gtaaattagg tttagtaaaa cattatagta ccacctggtc atgggttggt gtagatggca
4021 atgaaagaat aggtagatca cgtatgcttt taagttttac ttcaaacagc actagatcac
4081 agtatgttaa aattatgaag ctccctaaag gtgtggaatg gtcttttggt aattttgata
4141 agctttTAAca ttttgctaac atactaacgg tgcttgcact actaacacat TAAtcttta
         <- E2 end                                L2 orf start ->
4201 acatttttat attgcttttt tatttttata taATGgtgcg tgctcgcaga acaaagcgag
                                    L2 cds ->
4261 attctgtaac aaacatttac agaacctgca aacaggcagg aaactgtcct ccggatgttg
4321 ttAATAAAgt ggaacaaact acaattgctg accaaatatt aaaatttggc agcactggtg
signal ->
4381 tgtttttttgg tggtttggga ataggtacag gccgtggtac cggtggcagt actggctatg
4441 tacctatagg tgaaggccca gcaatacgtg ttggggggcac tccaagtgtt gttcgtccag
4501 gtatactccc tgaggctatt ggtccggcgg atatcattcc tattgatact gtcaatccaa
4561 ttgatccaaa tgcatcatct gtggtcccac tcactgacac aggacctgat ttgctacctg
4621 ggacaattga gactattgca gaagtgaacc ctgccccaga tattcctaga gttgacacat
4681 ctgttgtcac aacaagcaga ggctccagtg ctgtattgga ggttgcctct gaacccacac
4741 cacccactcg caccagaatt ccagaacac agtaccataa tccctctttt caaatattaa
4801 ctgaatctac accctctttg ggagaatctg cattaactga tcatgttgtt gttactagtg
4861 gttctggtgg tcaaccaata ggtggagtta caccagttga aatagaatta caagaacttc
4921 ctagcagata tactttttgaa atagaggaac ctaccaccac aagacgctct agtaccccac
4981 tacgcaacat cacacaagct gtaggaaatt aagaagatc actatataat aggcgactta
5041 ctcaacaagt aaatgtccag gatccattat cttacaaca gccctcacgt ttagttcgct
```

```
5101 ttgcctttga taatcctgtg tttgaagaag aagttacaca aatatttgaa agggacgtag
5161 cagctgtaga agaacctcca gacagagact ttttagatat agcaaaatta agccgccctc
5221 tttactctga aacaccacag ggatatgtca gggtaagccg cttaggtaat agggcttcta
5281 ttagaacacg tagtggagct acagtagggg ctcaagtgca tttttataca gatcttagca
5341 caatcgatgc agaggagtct atagagttat cactattagg ggaacattct ggtgatgcta
5401 ctattgtcca aggcccagta gaaagctcat ttgtagattt aaatgttcag gaactgcctc
5461 aagtaataga agtagaccca gaacctactt tccactctga tgatttgcta ctggatgagc
5521 aaaatgaaga ttttctggc tcccagttag tttatggtag tggcaggcgt tctaccacat
5581 ttactgtacc ccgcttctct actcccagat ctgatacctt ttatgtacaa gatttggaag
5641 gttatgctgt gtcatatcct gaacgaagga attatccaga aattatttat cctcaacccg
5701 atttgccaac tgtaataatt catactgcag atacctctgg ggacttctat ttacatccaa
5761 gccttcgcag gcgaaaacgt aaacgcactt atttaTGAta tttctttcag ATGacctcgc
                         L1 orf start ->    L1 cds ->
                                   <- L2 end
5821 tatggttacc tgcaactggt aaggtatatc taccaccttc aacacctgtg gcaagggtac
5881 aaaagcacgga tgaatacatt cagaggacag acatctacta tcatgctaat agtgatcgat
5941 tgttaactgt aggacatcca tattttgatg tgagagatac agcagacaat tctaaaattt
6001 tagtaccaaa ggtttcaggt aatcaatatc gagcctttag attactatta ccagatccca
6061 acagatttgc actagtagat atgaatatat ataacccaga aaaggaaaga ttagtatggg
6121 cctgtagagg cttagaaatt ggtcgtggcc agcctttagg tgttggtaca acaggacatc
6181 cattgtttaa caaagtcaaa gatactgaaa atgctaataa ctatatagta acttctaaag
6241 atgatagaca ggatacttca tttgaccct a aacaggtaca aatgtttatc ataggttgta
6301 ctccttgtat gggtgagtac tgggacgctg ctaaaccttg tgatgcagat gctggtcagg
6361 gtaaatgccc tccattagaa ttaatcaatt cagttataca agatggtgat atgattgata
6421 taggtttTgg taatatcaat aataagacat tatctgttaa cagatctgat gtcagtttgg
6481 atatagtaaa tgacatttgc aagtatcctg atttTttaaa gatggcaaat gacatatatg
6541 gggatgctgg tttcttctat gctagacgtg aacaatgtta tgccaggcac ttctttgtta
6601 gaggtggtaa tgtaggggat gcgataccca atactgctgt aggtcaggat aacaattaca
6661 tattacctgc agcaagtcaa caggcccaaa atactcttgg cagctccatc tatttcccta
6721 ccgtcagtgg ctctttggta tctactgatg cgcagctatt caatagACCT TTTTGGTTac
                                                      -> E2 bind
6781 aaagagcaca gggtcacaac aatggaattt gctgggagaa tcagcttttt ataacagtgg
6841 ctgataatac cagaaatacc aattttacta ttagtgtaag tacgdatggc cagacaccta
6901 cagaatatga cagtaccaag gttagagaat ttttaagaca tgtagaggaa tatgaaattt
6961 caattatatt acaattgtgt aaggtacctt tagaaccgga agtcctggca caaatcaatg
7021 ctatgaattc ttctatattg gaaaattggc aattgggatt tgttcctacc cctgataatc
7081 ctatacatga cacatatagg tatcttacat cacaggcaac acgatgccct gacaaacaac
7141 ctgctccaga aaggaaagat ccatatgagc agtataactt ttggactgta gatttaacag
7201 aaaaactgtc tttggatttg gatcaatatt ctttaggaag aaagtttta tttcaagctg
7261 ggctacaacg ggcttctaga gtgtctaaat cctctgctgc tagagcttcc acacggggta
7321 ttaaacgaaa acggagaTGA CCGTTTTCGG Ttgctgggtc ttataataaa atattttata
                      <- L1 end
                      -> E2 bind
7381 aactgttttg gtatgtgagg catgtttTaa ccgagttcgt gactaagatt gattaaccca
7441 cctgcaACCG CACCCGGTta atcagattat aaaggtgcgc cggtgttcac ctctggctac
              -> E2 bind
7501 ttggcagtta caagttcacc tctgccagaa gtgtgttttt gccaagacat ttgccaagtA
                                                         E2 bind ->
```